

Ethics can boost science

A report from a European ethics committee gives a valuable summary of the issues surrounding stem-cell research. Debates over therapeutic cloning should not distract attention from central ethical concerns and alternative stem-cell techniques.

On the trail of the latest Holy Grail, researchers can sometimes lose sight of the wider issues and of alternative avenues. Scientists lobbying passionately for research on human embryonic stem cells, and in particular on 'therapeutic cloning', should pay heed. Cold water is thrown on therapeutic cloning this week by a report from a group of 12 wise men and women, convened to advise the European Union (EU) on bioethics (see page 277). It concludes that the creation of embryos by somatic-cell nuclear transfer for research on stem-cell therapy would be "premature". Analysis of the legitimacy of that assertion could itself fill pages, and it is this aspect that will no doubt claim public attention, dealing as it does with the explosive cocktail of human cloning and embryo research.

Therapeutic cloning has potential advantages. In this technique, somatic nuclei from a patient's body — nuclei from non-reproductive cells — are fused with donated eggs (oocytes). Embryos are thereby produced from which pluripotent stem cells — those able to differentiate into any body-cell type, and genetically identical to the patient — could be generated. Cells or tissues derived from such embryonic stem cells for transplantation would not be rejected by the patient's body. The question of whether therapeutic cloning should be allowed is currently generating much heat in Britain, and elsewhere. But that is only one part of the vast field of regenerative medicine being opened up by stem cells. There is a danger that the issue of therapeutic cloning will distract politicians and the public from the wider issues and opportunities offered by embryonic stem cells for both therapeutics and basic cell biology.

The real interest of the report is in disentangling the broader range of issues. Britain is already at the far end of the moral spectrum, being one of the few countries in the world to authorize the 'utilitarian' deliberate creation of embryos for research, a practice that runs against the European Convention on Human Rights and Biomedicine. In most of Europe, and elsewhere, the fundamental question is not whether the cloning of human embryos to produce stem cells should be allowed, but whether the collection of stem cells from human embryos should be allowed at all. And it is here, in the exploration of the use of basic embryonic stem cells, rather than in somatic nuclear transfer, that the scientific promise is greatest.

Promise

Although countries opposed in principle to embryo research, such as hard-line Germany or Ireland, are unlikely to waver, the huge therapeutic promise of embryonic stem cells is likely to lead to changes in the law in countries such as France. There, a ban on human embryo research is likely to be lifted to allow embryonic stem cells to be collected, but only from the 35,000 frozen surplus embryos left over from attempts at *in vitro* fertilization. One need not be a Jesuit theologian to realize that the ethical argument for this is overwhelming: if such embryos are to be destroyed anyway, why not use them to relieve suffering? Obtaining consensus on whether human embryo research should be allowed is impossible, given the diversity of views. The EU

report is therefore wise to leave this question up to national legislatures, rather than seeking to impose Europe-wide rules on EU research funding.

But the ethical controversy over the use of human embryos as a supply of embryonic stem cells has perhaps done science a favour. Three years ago, before human embryonic stem cells had been cultured, xenotransplantation was being pursued as the solution to cell and organ transplants (see *Nature* 391, 325; 1998). The breakthrough in 1998 in culturing human embryonic stem cells has now radically shifted the focus of research. In turn, whereas the use of such cells has since become the obvious way forward, the controversy over the use of embryos is now encouraging research into 'alternative' sources that might otherwise have been ignored: adult stem cells.

Use of adult cells

Stem cells from adult tissues such as bone are currently of limited use in that they do not grow as easily as embryonic stem cells, or differentiate into as many cell types. But are adult stem cells irretrievably committed to differentiating into a limited number of cell types? The past year has seen reports of blood stem cells generating muscle tissue and muscle stem cells generating blood cells. Even one year ago, few would have believed such feats possible. Brain cells, once thought irreplaceable, can now be regenerated. Might all types of brain cells be regenerated from an adult human stem cell? Why not?

Adult stem cells still appear of limited use because they cannot proliferate indefinitely. But who says they are not capable of doing so? Differentiation is a complex process, determined by the switching on and off of thousands of genes and the production of millions of proteins. The intellectual appeal of human embryonic stem cells, with their ability to divide indefinitely and potentially generate all types of cells, should not be allowed to lead to a neglect of other avenues of research, particularly given the ethical issues involved. A case-by-case approach will probably be needed — in mice, the entire blood system can now be regenerated from a single adult blood stem cell.

A cynical view of a bioethics committee is that it simply puts a brake on the introduction of new technologies until public acceptance is inevitably in place. In the EU report, ethical concerns are not only doing science a service by providing a bridge to society's legitimate concern about issues such as the rights of the human embryo, but they are also giving science an opportunity to stand back and think of alternative approaches, rather than putting all of its oocytes in one basket.

In Europe, the starkly contrasting views on embryo research are similarly having the beneficial effect of forcing the continent to debate and come to terms with fundamental ethical and scientific questions. US politicians would perhaps do well to take note of the debate in Europe. By comparison, the current US legal situation — with moral restrictions in the public sector, and almost anything allowed in the private sector — is both hypocritical and ethically incoherent. ■





Sample delights

Bioprospecting in Mexico grinds to a halt
p278



Gradual thaw

Elections cast US global-warming policy adrift
p279



Monkey business

Question mark remains over fate of Japanese macaques
p280



Shut down

CERN's hunt for the Higgs boson is put on hold
p281

European panel rejects creation of human embryos for research

David Dickson, London

The debate over whether scientists should be allowed to create human embryos for use in stem-cell research is set to take a dramatic twist this week. A report giving top-level advice to the European Commission will rule that approving such practices would be "premature".

"The creation of embryos for the sole purpose of research raises serious concerns since it represents a further step in the instrumentalization of human life," says the report. It goes on to describe the hopes of regenerative medicine as "still very speculative".

Although aimed primarily at guiding the commission on decisions about its own Framework research programme, the judgment is expected to have a significant impact on those European Union member states, including France and Britain, that are currently debating the issue.

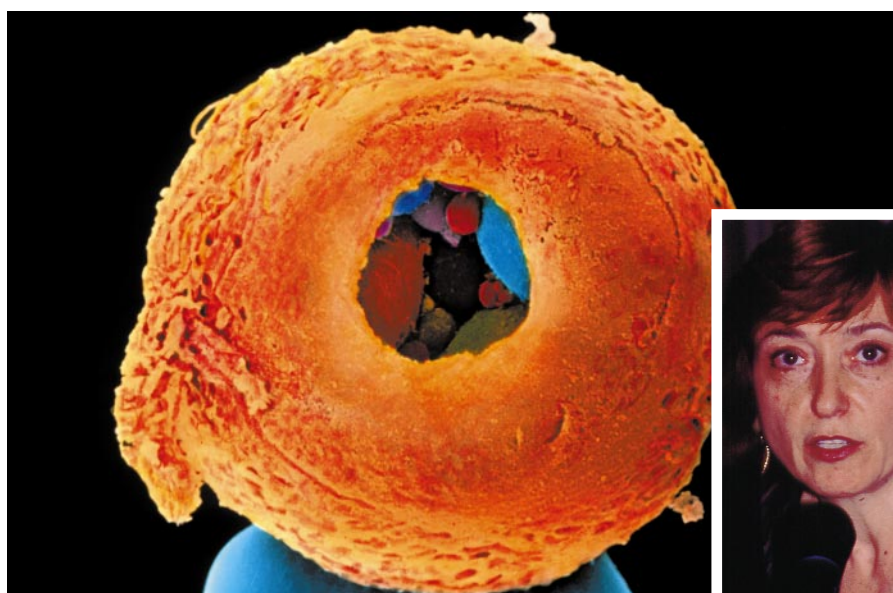
The report comes just days before the British parliament is likely to debate a recommendation from the government's chief medical officer, Liam Donaldson, that such research should be allowed to proceed (see *Nature* 406, 815; 2000).

In the United States, meanwhile, the National Institutes of Health's plan to allow research on stem cells—but not their extraction for that purpose—has been left hanging in the balance by the inconclusive results of last week's presidential election (see page 279).

The new report was prepared by the 12-member European Group on Ethics in Science and New Technologies (EGE), chaired by French high-court judge Nöelle Lenoir. It was due to be presented to the European Union's Council of Ministers this week.

Romano Prodi, president of the European Commission, requested the report to provide guidance on what has become a hot potato in Brussels. In September, the European Parliament narrowly approved a resolution declaring its opposition to the 'therapeutic cloning' of human embryos—which it said was no different from reproductive cloning—and demanding criminal penalties for any infringement.

Produced after a series of hearings with



Handle with care: Nöelle Lenoir's report recommends 'prudence' for stem cells.



scientific and religious witnesses, the EGE's report acknowledges both the "scientific justification" and the potential medical value of somatic-cell nuclear transfer. But it recommends "prudence" and a precautionary approach, concluding that "at present the creation of embryos for somatic cell transfer would be premature".

Lenoir says that the conclusion represents a compromise between three different positions expressed on her committee: that all embryo research in member states should be outlawed (as it is in Germany, for example); that it should be permitted, but that the creation of embryos for research purposes should not be; and that the latter should be allowed under carefully regulated circumstances, as is being proposed in Britain.

She says that the panel's unanimous decision that allowing such research to proceed would be premature represents "a pragmatic approach" reflecting a need to find a consensus as a basis on which European research can be funded. Lenoir also emphasizes that the group did not want to cut off future options, and "is not trying to interfere in any way with national legislation".

The British government's position has

already been complicated by the heavy defeat two weeks ago of a private member's bill supporting such research submitted by Evan Harris, Liberal Democrat member of parliament for Oxford, West. As they have done in the United States, opponents claimed that the creation of embryos for research purposes was unethical, and that therapeutic cloning is the first step on a slippery slope to reproductive cloning.

Some members of parliament are claiming that they voted against the bill primarily because of the strictly limited time for debate allowed on such private bills. And the scientific community has launched a vigorous lobbying effort, spearheaded by the Royal Society, to support the implementation of legislation allowing embryonic stem-cell research.

But other observers say that the heavy defeat of Harris's bill will have sent a warning to the government that the free vote it has promised on the issue could prove an embarrassing failure, particularly in a period prior to a general election. They add that delaying a vote until after the election could increase the chances that legislation permitting stem-cell research is approved.



Bounty hunters? Some groups claim that bioprospecting exploits indigenous people.

Political uncertainty halts bioprospecting in Mexico

Rex Dalton, Mexico City

International bioprospecting in Mexico has ground to a halt amid uncertainty about the new government's approach to the collection and export of plant samples, and a vocal campaign that has branded the activity as exploitative.

Two US-funded teams — each involving scientists from US and Mexican universities — have stopped searching for specimens because they cannot secure the appropriate permits from the Mexican government. A third effort involving scientists at the National Autonomous University of Mexico (UNAM) and Diversa, a Californian biotech firm, was halted earlier in the year.

The slowdown comes as Mexico undergoes a historic change of government, with Vicente Fox ending 80 years of rule by the Institutional Revolutionary Party when he becomes president next month. At the same time, the Rural Advancement Foundation International (RAFI), an advocacy organization based in Canada, has mounted a public campaign alleging that bioprospectors are exploiting indigenous people.

The blocking of the projects has dismayed US and Mexican scientists alike. "It is very unfortunate," says Joshua Rosenthal of the Fogarty International Center at the US National Institutes of Health (NIH), which funds the two projects. "What is required now is some clarity from the new government in Mexico."

Fernando Tudela, chief of staff for the Mexican ministry for the environment, natural resources and fisheries, says: "We share the frustration of the scientists. But Mexico lacks a legal framework for bioprospecting. I would not advise undertaking one of these projects now."

Tudela, who loses his post when the new government takes over, adds: "My advice to the next administration is to carefully draft a

plan for broad discussions of rules and guidelines for bioprospecting."

Brent Berlin, a University of Georgia ethnobiologist who has worked in southern Mexico for 30 years and whose project in Chiapas was recently blocked, says that "paralysis" has enveloped bioprospecting in Mexico. "The work is going to happen, but at a slower pace," says Berlin. His project involved interviewing Mayans in the highlands near Mexico's southern border, identifying the substances their healers use for therapies, and collecting plants for analysis. Identified active compounds then might be used for drugs or agricultural products.

Before the current impasse, Mexico had issued a handful of permits for scientific collections under the international Convention on Biological Diversity's guidelines for universities, biotechnology companies and indigenous people. One such permit allowed El Colegio de la Frontera Sur (ECOSUR) in Chiapas, working with Berlin, to collect about 5,600 plant specimens.

But RAFI and allies have branded the project "biopiracy", alleging inadequate Mayan participation. Berlin and ECOSUR scientists say that RAFI has maligned them unfairly. But when ECOSUR applied to Mexico's National Institute of Ecology for a permit to collect plant samples for bioassays for commercial products in September, after Fox's election, the institute made demands that the team found impossible to meet.

At the same time, US officials became concerned about another bioprospecting project, involving the University of Arizona and UNAM's Botanical Garden in Mexico City. Arizona had a contract with the pharmaceutical corporation American Home Products to develop any discoveries from the collaboration. But after problems arose with the Mayan project permit, the NIH's Rosenthal advised UNAM to halt the collections.

Online naming of species opens digital age for taxonomy

Henry Gee, London

The ancient art of taxonomy — the classification of organisms — will take a leap into the twenty-first century this month with the first publication of zoological names in an online journal. The move has radical implications for standards of access to archive material that taxonomists have accepted for almost 250 years.

The recipients of the new names are three species of fossil protists called foraminifera. They appear in a paper by David B. Scott of Dalhousie University in Halifax, Nova Scotia, and colleagues, to be published in *Palaeontologia Electronica*.

Taxonomy relies on access to archive material stretching back decades or even centuries. By convention, names created after 1758 — when the tenth edition of Linnaeus' *Systema Naturae* was published — must use the binomial nomenclature established in that work. They must also follow the principle of priority of publication enshrined in the International Code of Zoological Nomenclature — known in the trade simply as 'the Code'.

Until recently, the Code gave no quarter to the electronic age — to be valid, formal names had to be printed on paper. But its 1999 edition conceded that names created after 1999 may be valid if published "by a method that does not employ printing on paper", provided that numerous identical and durable copies are obtainable and that copies are deposited in at least five named libraries.

As well as library distribution, *Palaeontologia Electronica* will make 200 CD-ROM copies of the issue containing the paper available for a fee. The new name will become valid when the disks, are distributed, although it is intended that both versions will be issued simultaneously.

Despite the concession, the Code still emphasizes the "desirability of works on paper". The reason is clear — even if CD-ROMs were to last for centuries, hardware developments might make it difficult to find machines that can read them.

But Norman MacLeod of the Natural History Museum in London, who edits *Palaeontologia Electronica*, believes that electronic taxonomy will be "very much the rule" within a decade. "I would be severely depressed if in ten years I walk into an office and find a series of bound volumes of *PE* printouts," he says.

♦ <http://palaeo-electronica.org>

Election impasse leaves science in the dark

Tony Reichhardt and Paul Smaglik, Washington

The deadlocked race for the US presidency, together with uncertainty over the extent of Republican control in the Senate, has left a huge question mark over who will be calling the shots in science policy in January.

The traditional scramble for key positions in the new administration — such as that of the president's science adviser, and the vacant directorship of the National Institutes of Health (NIH) — was put on hold as the campaigns of Al Gore and George W. Bush concentrated on manoeuvring to determine the final outcome of the election.

But in the House of Representatives, where Republicans will keep control by a small, as yet unspecified, margin, some committee posts are already being claimed. Sherwood Boehlert (Republican, New York), a moderate republican with a strong interest in environmental protection is widely expected to become chair of the Science Committee, which has jurisdiction over most non-biomedical research programmes. Its current chairman, Jim Sensenbrenner (Republican, Wisconsin), hopes to move on to chair the Judiciary Committee.

Most leadership positions in the House will change as a result of a new rule that requires committee chairpersons who took their position six years ago to step down. "There will be a lot of musical chairs," predicts Howard Silver, who tracks science legis-

lation as executive director of the Consortium of Social Science Associations.

Some of the most influential proponents of the doubling of the NIH's budget are either retiring or considering stepping away from the fray in the new Congress. And their replacements may lack either their visibility or their zeal, lobbyists for biomedicine say.

John Porter (Republican, Illinois), chair of the House subcommittee that funds the NIH and an ally of biomedical research, is retiring and Ralph Regula (Republican, Ohio) is the leading candidate to replace him. Other contenders are Henry Bonilla (Republican, Texas) and Ernest Istook (Republican, Oklahoma).



Citing dissatisfaction that the NIH budget did not come to a vote before the 2000 election, Senator Arlen Specter (Republican, Pennsylvania) has threatened to give up the chair of the Senate appropriations subcommittee that funds the agency. Specter has strongly supported the bid to double the NIH's budget in five years and is also a keen proponent of stem-cell research.

One possible successor, Senator Slade Gorton (Republican, Washington), is involved in the only Senate race that had not been called one week after the 7 November elections. The Republicans had won 50 seats and the Democrats 49, including that of Joe Lieberman (Democrat, Connecticut), Al Gore's running mate. Because Lieberman cannot sit in the Senate and be vice-president, Republicans are certain to maintain control of the chamber, albeit by a razor-thin margin.

With such an even split in numbers, bipartisanship and moderation will be needed for any legislation to survive in the 107th Congress. As Silver puts it: "Anything that looks radical is not going to pass." That should put basic research, which has strong bipartisan support, in a strong position.

But the body politic remains deeply divided along party lines on at least three science-related issues: global warming (see below), embryonic stem-cell research and national missile defence.

Climate talks face uncertainty over US strategy

The question of who will be in the White House come January will cast a heavy pall over this week's climate treaty talks in The Hague. And, ironically, some observers think that George W. Bush would be better placed than Al Gore to break the political stalemate in the United States on global warming.

At first glance, a Bush presidency would seem to be the environmentalists' worst nightmare. He and running mate Dick Cheney are both oilmen, and their energy strategy calls for opening up part of the protected Arctic wilderness to drilling.

Yet Bush is not as combative on these issues as many members of the Republican-controlled Congress, and he understands that a modern politician must at least sound like an environmentalist. Early in his campaign, Bush expressed doubt about the science behind climate-change predictions, but later recanted, stating: "I believe there is global warming."

Although he opposes the Kyoto guidelines as "ineffective, inadequate and unfair to America", so do many other US politicians in both parties. The 1997 protocol, which called

for sharp reductions in greenhouse-gas emissions by 2008, was dead on arrival in Congress, and the body has not stirred since.

That leads Henry Jacoby, an environmental economist at the Massachusetts Institute of Technology's Sloan School of Management, to contend that "if you wanted to see progress on the climate issue, you would have voted for Bush. Gore has so much opposition, and is so

locked in to [the Kyoto] strategy" that it may not be possible for him to break the impasse.

The US delegation to the Sixth Conference of Parties to the United Nations Framework Convention on Climate Change, led by under-secretary of state Frank Loy, will represent the Clinton-Gore administration. As of early this week, Bush was not planning to send a representative to the talks.

T. R.



Ice breaker: Bush wants to drill for oil in the Arctic, but could he also be a positive force for change?

Row over fate of endangered monkeys

David Cyranoski, Tokyo

Japanese researchers and animal-welfare activists are at loggerheads — again — over the treatment of wild monkeys that are captured, sold to laboratories and used for research in neuroscience and physiology.

The country's Environment Agency is preparing new guidelines on the treatment of Japanese macaques, for implementation in the new year. But neither researchers nor activists are happy with the agency's draft, which is open for public comment until 17 November.

The use of macaques in research is a long-running bone of contention in Japan (see *Nature* 393, 404; 1998). Although, as an endangered species, they are not on the list of animals that can be hunted, the monkeys can wreak havoc in farming communities, and



Look out: the law aimed at protecting Japan's threatened macaques is under discussion again.

local governments grant permission to 'remove' them when a damage claim is submitted. More than 53,000 have been collected in this way since 1993.

Problems arise after the monkeys are caught. Legally, those who capture monkeys

must report only how many they have taken. Monkeys are either killed, put in a preservation park, or sent to research facilities.

According to the animal-protection group ALIVE (All Life in a Viable Environment), between 1,000 and 2,000 macaques end up in laboratories each year. In one case, a local government sent monkeys to a nearby medical school, but claimed to have put them in a preservation park. The transfer itself is not illegal, but misrepresenting the destination is.

"Although the exact legal framework to punish such violations is unclear, such cover-ups are an acknowledgment of law-breaking," says ALIVE spokesman Fusako Nogami, adding that animals taken in this way should be off-limits for research.

The Japanese Society for Primate Research says that removed monkeys should be put down rather than used for medical research. But other scientific societies want the animals to be used in research. They claim that researchers follow strict guidelines on the care and testing of animals, in line with international standards.

The Environment Agency's draft is not legally binding, but it is likely to influence the application of binding laws, and is therefore being hotly debated.

The draft's most controversial phrase states that the monkeys may be used for "research on the ecology or health of the monkeys themselves". Researchers fear that such a suggestion, along with pressure from animal-rights groups, will persuade judges to restrict research to this single case. They say that another proposal, requiring applications for removal to state the monkeys' fate after capture, would discourage local governments from giving macaques to research.

A representative for the Japan Neuroscience Society predicts that the supply of macaques would fall, raising their price from an affordable ¥150,000 (US\$1,400) to a prohibitive ¥1 million. This, he says, would make research impossible for all but a few of the 40 or so laboratories currently working with monkeys.

The representative says that the Environment Agency's proposal underestimates the importance of the research, without which, he says, "we would know nothing about the human brain". He considers that the guidelines should endorse the use of macaques in research.

Nogami counters that the removal of macaques is used to justify what has become a systematic procurement of monkeys for research. Since the World Conservation Union lists Japanese macaques as endangered, she adds, the guidelines should prohibit their use and lay out clearer rules for the punishment of transgressors.

Fake finds reveal critical deficiency

Last week's disclosure of extensive fraud by one of Japan's leading archaeologists has led to renewed soul-searching about how much the nation's more prominent scientists are allowed to escape criticism by their peers.

The self-examination began after journalists from *Mainichi Shimbun*, one of Japan's leading newspapers, twice filmed Shinichi Fujimura burying stone tools he had collected in previous archaeological digs, to unearth them again as prehistoric records. Archaeology is a national obsession in Japan, and Fujimura's 'discoveries' had often made front-page news.

Fujimura — who, according to critics, had no advanced degree and little scientific training — was last week sacked from his post as senior director of the Tohoku Paleolithic Institute. The episode has raised questions about a scientific culture that allowed Fujimura's work to continue, and to be taken seriously, despite suspicions that, it

now emerges, stretched back many years.

"In Japan, it is hard to criticize people directly, especially those in established positions, because the critique is taken as a personal attack," explains Hisao Baba, a palaeoanthropologist at the National Science Museum in Tokyo. He hopes this episode will help usher in a culture of research in which criticism is accepted at a scientific rather than a personal level.

The problem has been widely noted in Japan. A visiting physicist, for example, confided that he felt pressure to tone down research that overturned a leading Japanese physicist's findings as it would have been considered a "personal insult".

According to Shizuo Oda of the Tokyo Metropolitan Division of Cultural Affairs, this soft-peddling can make researchers as limited in their views as "a frog in a well that is unaware of the ocean". Academics may not care about other researchers' work, or about having their own work peer-reviewed. This is especially dangerous in an archaeological community where, Oda says, intuition is sometimes valued over facts.

Fujimura claims that all the work he did before the filmed forgery was authentic. But most Japanese archaeologists are waiting for the rest of his findings to be discredited. Several archaeologists and anthropologists are now saying that Fujimura's discoveries did not mesh with other information about the sites. His discoveries suggested that the Japanese Stone Age stretched back as far as 700,000 years ago. But Oda has long held that Japan before about 50,000 years ago. **D. C.**



Here's one I found earlier: archaeologist Shinichi Fujimura with one of his finds.

Promise of Higgs fails to save CERN collider

Alison Abbott, Munich

After eleven years and a nail-biting three-month reprieve, the death knell for the Large Electron-Positron collider (LEP) in Geneva has finally been sounded.

Last week, Luciano Maiani, director-general of CERN, the European Laboratory for Particle Physics, rejected requests to keep the LEP running for another year, ruling that it should be dismantled early in the new year as planned.

The drama over the collider's future centred on the possible sighting by the LEP's detectors of the subatomic particle known as the Higgs boson (see *Nature* 408, 10; 2000). Showing that Higgs exists would bolster the Standard Model of particles and forces, and fulfil one of CERN's major missions.

The LEP had been granted a one-month extension to gather more data in support of the initial observation. As the probability that Higgs really had been seen fell at first, only to rise in the past few weeks, pressure grew on CERN's management to keep the LEP running for another year.

But that would have cost SFr100 million (US\$57 million) and delayed construction of the Large Hadron Collider (LHC), which will be powerful enough to confirm the Higgs if, as the LEP experiments suggest, it has a mass of 115 giga-electron volts.

High-energy physicists were unable to



Switched off: accelerating cavities from the soon to be dismantled LEP collider in Geneva.

reach a consensus on whether the risk should be taken. The LEP scientific committee comprising LEP researchers, including some with experiments planned for the LHC, split

50–50 on whether to recommend the extension to CERN's research board. The board, whose job is to balance the whole of CERN's programme, also failed to reach a consensus. Even appeals to CERN's external Scientific Policy Committee drew an inconclusive response.

That left Maiani to bear responsibility for the decision. He took it, he says, not on a financial basis, "but on the basis of optimizing the overall scientific return of the laboratory".

"The conclusive discovery of Higgs within a year was not guaranteed and I thought it too risky to delay the LHC and distract the scientists who are going to run its experiments," he says. The LHC will, Maiani points out, let physicists study the properties of Higgs as well as just identifying it.

Roger Cashmore, CERN's research director, adds: "The best way forward for CERN is to proceed with the LHC as fast as possible. To change a good research programme you need a very strong set of arguments, which the committees did not supply."

But Chris Tully, one of the CERN physicists who worked on the Higgs search data, says the decision is "tragic for CERN". He believes CERN management has underplayed the strength of the evidence that the Higgs has been sighted.

♦ <http://public.web.cern.ch/Public>

Medical institute opens amid hopes of Kansas rebirth

Paul Smaglik, Kansas City

The people of Kansas City, Missouri, are hoping that the official opening this month of the \$200 million Stowers Institute for Medical Research will provide a springboard that will eventually transform the midwestern city into one of the nation's top life-science centres.

The institute — built by a married couple of wealthy cancer survivors who also donated its initial endowment of \$515 million, which could rise to \$2 billion — reflects the growing influence of large-scale philanthropy on biomedical research in the United States.

The institute will not on its own revitalize Kansas City's research climate, as it will initially hire only four to six investigators a year, but it has already leveraged other activity aimed at doing just that.

Jim and Virginia Stowers, who made their fortune in the mutual-funds business in the city, say they ignored advice to spend their fortune on an institute in a more established centre of biomedical research on America's east or west coasts.

The advent of the institute has prompted the city to form a coalition of businesses, hospitals and universities called the Kansas City Area Life Sciences Institute (LSI). The coalition hopes that the high profile the new institute will bring to the area should help it meet its target of raising \$30 million a year locally over ten years. The LSI estimates that Kansas City's annual biomedical research effort, supported by the National Institutes of Health and other sources, could by then have expanded from \$100 million to \$500 million.

Bill Neaves, the president of the Stowers Institute, says he hopes that the LSI will help him to recruit top scientists. Before hiring anyone, he introduces them to scientists working at LSI coalition institutions, such as the Midwest Research Institute and the University of Missouri in Kansas City. Dual appointments and the chance to use facilities outside the Stowers Institute will attract talent, he believes.

Neaves has proceeded cautiously, under the guidance of Doug Melton, chair of molecular and



Room for growth: Neaves (inset) hopes the new institute will help attract talent to the area.

cellular biology at Harvard University. Neaves has so far hired four principal investigators, led by Robb Krumlauf, who previously ran the division of developmental neurobiology at Britain's National Institute for Medical Research at Mill Hill, London. "There's a danger in going too slow, but there is perhaps a greater danger of going too fast," says Neaves.

♦ <http://www.stowers-institute.org>

Australia appoints outsider to lead research agency

Sydney In a surprise move, Australia is bypassing local candidates and bringing Geoff Garrett from South Africa to lead its largest



Geoff Garrett, new head of the CSIRO.

research agency, the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Garrett, a 52-year-old former metallurgist, was nearing the end of his term as head of South Africa's main research agency, the CSIR. He has a strong track record in attracting external funds for the South

African agency, where they make up 60% of the budget.

The position at the CSIRO — the most important in Australian science, controlling a budget of A\$770 million (US\$400 million) — became vacant when Malcolm McIntosh died in February. There has since been a stalling of institutional momentum at the agency, observers say, at a time when the

government has been struggling to articulate a policy for science.

An effort to raise the CSIRO's external income above its mandated level of 30% of its total budget has led to claims that it is being turned into a research and testing contractor for government and industry.

Europe's best networkers scoop Descartes prizes

Paris The European Union's new René Descartes science prize has awarded 600,000 euros (US\$517,000) to three trans-European research groups. Candidates were judged not only on their science but also on their skills in international networking.

The winners were chemists at the universities of Birmingham and Rennes, for developing methods for studying chemical reaction kinetics at very low temperatures; a team including Philips Research Laboratories in Eindhoven, the Netherlands, Eindhoven University of Technology, the University of Cambridge, Risø National Laboratory in Denmark, and the Universität Ulm in Germany, for research on plastic transistors; and groups at the University of Sussex, Italy's Consiglio Nazionale delle Ricerche in Pavia, the Erasmus University, Rotterdam, and the Centre National de la Recherche Scientifique in Strasbourg, for

discovering the XPD gene, which is involved in DNA repair and transcription.

Plight of cod demands drastic measures

London The International Council for the Exploration of the Sea has warned that cod stocks in UK waters will collapse unless drastic restrictions on fishing are introduced. Catches of North Sea cod have plummeted in recent years and the council says that the population is now "outside safe biological limits".

Environmentalists are warning of a collapse in stocks that could mirror the 1992 collapse in Newfoundland, Canada, that put 40,000 fishermen out of work. The European Union's fisheries council will set restrictions next month that may include the wholesale closure of some fishing grounds in 2001.

Pfizer faces competition after ruling on Viagra patent

London The High Court ruled last week that the pharmaceutical company Pfizer's UK patent blocking the development of anti-impotence drugs similar to Viagra was invalid. Other companies may now develop medicines based on compounds belonging to the same class of enzyme inhibitors as sildenafil citrate — Viagra's active ingredient.

The case was brought by Pfizer's US competitors Eli Lilly and ICOS. "We were confident that Pfizer's method of use patent was overly broad," said a spokesman for Eli Lilly. Pfizer, which is considering an appeal, said that the basic patent on Viagra will remain in force until 2013.

► <http://www.courtservice.gov.uk/judgments/viagra.htm>

Astronomer and ecologist take Balzan

Rome Italy's president Carlo Ciampi this week awarded the Balzan Prizes, worth SFr500,000 (US\$283,000) each, to Swiss astronomer Michel Mayor and Finnish ecologist Ilkka Hanski.

Mayor, director of the Geneva Observatory, was honoured for developing the instrumentation that led to the first detection of a planet outside our Solar System. Hanski won his award for his work on metapopulation dynamics, a field that examines patterns of extinction and migration in fragmented habitats.

Report slams anthropology book as 'fun-house falsity'

London The preliminary report of a team of anthropologists from the University of California at Santa Barbara (UCSB), has concluded

that allegations made against two scientists in a recent book appear to be "deliberately fraudulent".

The allegations of abuse against a remote South American people were made against Napoleon Chagnon and the late James Neel in journalist Patrick Tierney's book *Darkness in El Dorado* (see *Nature* 408, 137–138; 2000). But in the executive summary of an 11 November report, UCSB anthropology professor John Tooby writes: "Almost anywhere we scratched the surface, a massive tangle of fun-house falsity would erupt through."

In a statement, Bruce Alberts, president of the National Academy of Sciences, said: "Although [the book] gives the appearance of being well-researched, in many instances the author's conclusions are either contradicted or not supported by the references he cites."

► <http://national-academies.org/nas/eldorado>

Japanese to probe Tutankhamon's DNA

Tokyo Archaeologists at Waseda University in Tokyo have announced plans to analyse the DNA of King Tutankhamon. The researchers hope to reveal the lineage — and cause of death — of the Egyptian boy king, thought to have died 3,300 years ago, aged 18.

The mummified body was excavated in

1922, but the coffin has not been opened in 30 years. DNA from hair, bone and nail samples will be compared to that from two children of Amenhotep III, believed to have been Tutankhamon's father.

Affymetrix infringed Oxford patent, says US court

Washington A federal jury has ruled that the California-based biotechnology company Affymetrix's gene chips infringed a patent held by Oxford Gene Technology. Affymetrix had already lost a similar ruling in British courts (see *Nature* 404, 697; 2000).

The ruling may open the way for other companies to produce microarrays, because Oxford Gene Technology wants to license the technology. But Affymetrix plans to challenge Oxford's patent in the US courts.

Clarification On page 824 in the issue of 19 October 2000, a photograph of Max Planck flanked by two uniformed Nazis appears with the caption: "Close ties: Max Planck (centre) at the Kaiser Wilhelm Institute for Metal Research in 1935." This wording implies that Planck might have been a Nazi sympathizer. This was not the case — in fact Planck resigned his post as president of the Kaiser Wilhelm Institutes in 1937 in protest at the Nazis' treatment of Jewish scientists.

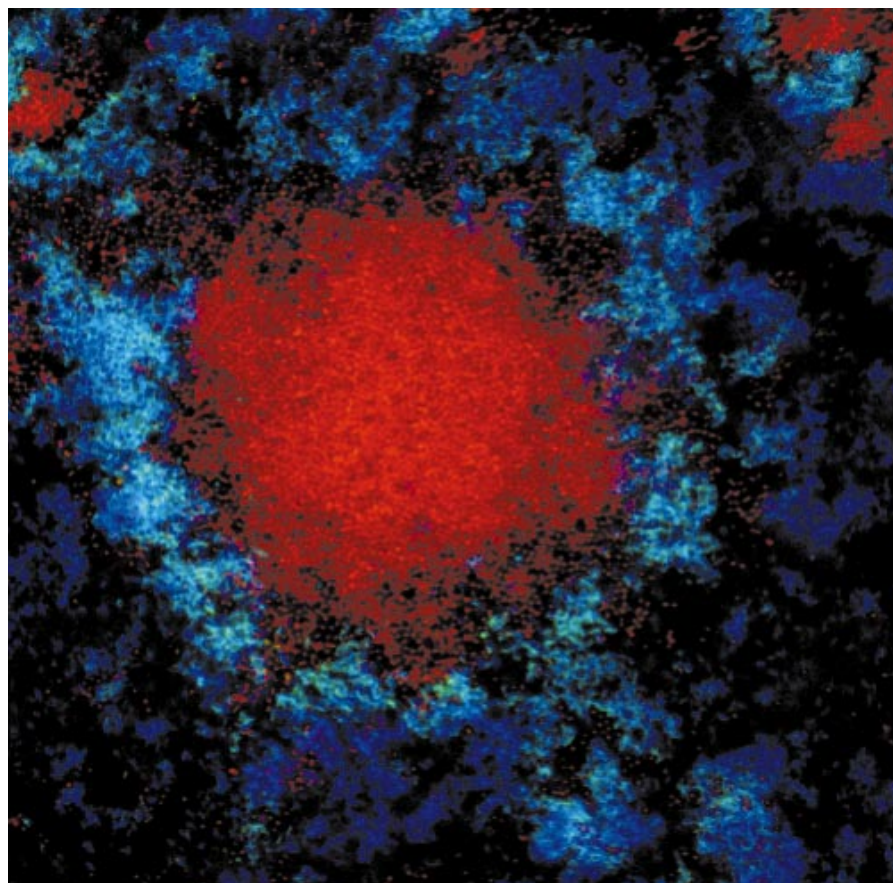
Slimebusters

Bacteria do not always simply float around — more often they grow on surfaces in mucilaginous communities called biofilms. Working out how to block their formation or dismantle them could help treat life-threatening infections, says Marina Chicurel.

Imagine yourself shrunk to the size of a bacterium. William Costerton, director of the Center for Biofilm Engineering at Montana State University in Bozeman, does so frequently. “If you found yourself in a biofilm, you’d be going along a channel full of water, like the canals in Venice, and up from the bottom of the channel, on either side, would be these slime towers,” he says. “The channels would be bringing in oxygen and nutrients, and removing waste. And within each building, so to speak, some of the bacteria would be cooperating with each other, making one compound and passing it along to the next. It’s at least as complicated as a tissue, and possibly as a city.”

Costerton was instrumental in alerting the world to the importance of these biofilms, coining the term in 1978¹. Today, microbiologists have realized that most species of bacteria adjust their biochemistry and behaviour to form slimy microbial cities when conditions allow. Biofilm-dwellers pump out sugar polymers called exopolysaccharides to create a complex, three-dimensional matrix, which typically comprises 85% of the volume of a biofilm. And within the past two years, researchers have made great strides in understanding the triggers that cause free-floating bacteria to form a biofilm and adjust to a sedentary existence.

From this vantage point, many possibilities are emerging, including the potential for designing antimicrobial drugs with fewer side effects that are less likely to promote resis-



In harmony: when a biofilm of *Pseudomonas putida* (blue) and *Acinetobacter* (red) grows on a medium containing benzyl alcohol, the two species can become tightly associated (blue/green). This is because, as *Acinetobacter* metabolizes the alcohol, it excretes benzoate, a major carbon source for *P. putida*.

tance. “Biofilms are a hot topic now because of the recognition that they are medically important,” says Roberto Kolter, a microbiologist at Harvard Medical School in Boston. Up to 60% of hospital-acquired infections involve biofilms², which contaminate implants and catheters. Biofilms also cause diseases ranging from the lung infections that can afflict cystic fibrosis patients to tooth decay^{3,4}. Add to this the fact that they cause billions of dollars’ worth of damage by coating equipment such as cooling systems, plus their role in bioreactors used for such tasks as manufacturing antibiotics, and it is easy to see why biofilm research has moved to the fore.

To create biofilms, says Kolter, microorganisms must be aware of environmental conditions — including the availability of nutrients and oxygen — and take a census of their neighbours, determining their density and characteristics. Many clues to the mechanisms of biofilm formation have come from screening mutants created using transposons — snippets of DNA that insert randomly into chromosomes, disrupting the genes into which they land. Researchers seed hundreds of tiny wells with individual mutants, allow them to attach, and then wash off the free-floating cells. Mutants that fail to leave their mark on the wells can then be studied to map their mutations, in the knowledge that the

affected genes are important for biofilm formation. These studies have shown that there is no universal set of genes involved. “What applies for *Vibrio cholerae* need not apply for *Pseudomonas aeruginosa*,” says Kolter. In addition, a microorganism can use different genetic pathways to build biofilms, depending on the environmental conditions.

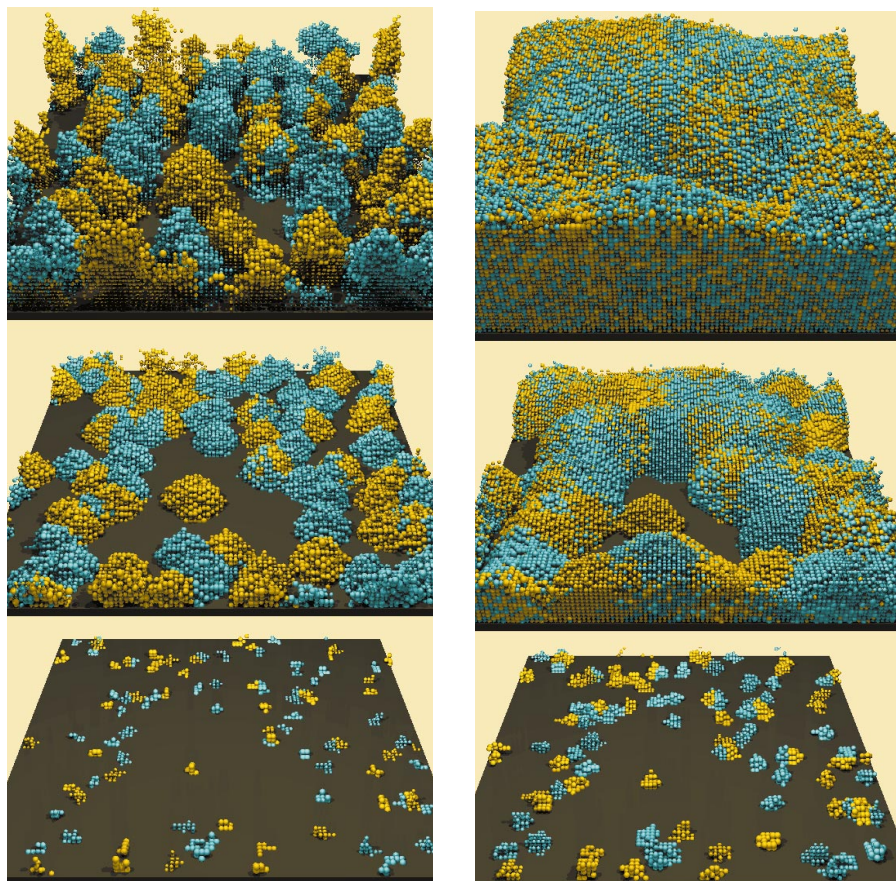
In several cases, the first steps of biofilm formation require genes involved in building force-generating appendages^{5,6}. *Pseudomo-*



Lethal layers: biofilms are thought to be behind fatal lung infections in cystic fibrosis patients.

SOREN MOLIN

WILL & DENI MCINTYRE/SPL



Virtual biofilms: computer simulations of the growth of a biofilm containing two species (blue and yellow). When nutrients are scarce and unevenly distributed (left), mushroom-like colonies form. A smooth and compact biofilm forms when the medium is rich in nutrients (right).

nas aeruginosa, for example, uses flagella to swim along surfaces, as if scanning for potential sites to colonize. The early settlers appear to move towards each other to form micro-colonies, using thin, retracting strands of protein called type IV pili to pull themselves across the surface^{7,8}. Other bacteria rely simply on cell division to begin the process⁹.

Once established on a surface, microbial colonists activate genes involved in producing the slimy matrix. Other genes appear to be involved in biofilm growth and in the adaptation to a communal lifestyle. Philippe Lejeune at the Lyon National Institute of Applied Science in Villeurbanne, France, has used the marker gene *lacZ* to track changes in gene expression as *Escherichia coli* forms a biofilm. His studies suggest that the expression of 38% of *E. coli*'s genes are altered¹⁰. But other groups report much more modest changes. In unpublished studies, researchers led by Peter Greenberg at the University of Iowa, working with colleagues at Harvard, have found that fewer than 200 of *P. aeruginosa*'s roughly 5,200 genes significantly change their expression in a mature biofilm.

Molecules called quorum-sensing signals help trigger and coordinate these changes. Bacteria constantly secrete low levels of these signals and sense them through receptors on their surfaces. But the receptors do not trigger

any behavioural changes until there are enough bacteria to allow the signals' concentrations to exceed a critical threshold. Once this occurs, bacteria respond by adopting communal behaviour, such as forming biofilms and, in the case of pathogenic bacteria, deploying virulence factors such as toxins^{11,12}. Illustrating the importance of quorum-sensing signals, Greenberg's team showed in 1998 that a *P. aeruginosa* mutant unable to produce a particular signal formed much thinner and more crowded biofilms¹³.

Exactly how quorum-sensing signals regulate biofilm formation remains a mystery, however. Several studies have revealed that their effects depend on environmental conditions and vary widely between species. In an effort to dissect the signals' effects, Greenberg's team has created a collection of mutants in a strain of *P. aeruginosa* incapable of secreting their own signals, but able to respond to synthetic quorum signals¹⁴. So far, they have found some 70 genes whose expression ramps up in the presence of the signals.

In addition to communicating with members of their own species, bacteria also conduct inter-species conversations. "It's not good enough to be able to only count yourself," says Bonnie Bassler of Princeton University in New Jersey. Bassler has discovered

that a family of genes involved in the synthesis of a particular quorum-sensing signal occur in a wide variety of bacteria¹⁵. In unpublished work, she has isolated the molecule and determined its chemical structure. The signal may form the basis for a bacterial Esperanto, she says, helping bacteria sense the presence of others. Although the mechanistic details remain to be discovered, Bassler suspects that bacteria may compare levels of this signal with those of their own species-specific signals to determine the proportion of their own kind in a growing biofilm.

Bacteria may also rely on signals for disassembling their communities. Kolter, for example, speculates that a lack of food may induce bacteria to produce signals that trigger an exodus from a biofilm.

Food for thought

But quorum-sensing signals are not the entire story. Søren Molin at the Technical University of Denmark in Lyngby has shown that biofilm formation can be shaped by food¹⁶. When fed chlorobiphenyls, species of *Pseudomonas* and *Burkholderia* form mixed biofilms, allowing *Pseudomonas* to feed off the metabolic products generated by *Burkholderia*. But when fed citrate, each species builds its own biofilm, with radically different structures. In addition, if the food source is switched, the bacteria rapidly restructure their biofilms. Molin's results suggest that the structure of biofilms can be explained by bacteria travelling up food gradients, without invoking signalling between cells¹⁶. Computer simulations by Cristian Picioreanu at the Delft University of Technology in the Netherlands suggest that, indeed, nutrients are not evenly distributed within biofilms and may help shape their structure¹⁷.

Work by other researchers has revealed that oxygen is unequally distributed within a biofilm — with bacteria in the deepest reaches often living in almost completely anaerobic conditions¹⁸. "Even small domains of a community may have a composition around them that is quite different from that just a few micrometres away," says Molin. He recently illustrated just how different the lives of individuals in the same biofilm can be by tagging bacteria with fluorescent proteins, and observing their movements⁹. Instead of finding only sedentary bacteria, Molin saw some bacteria swimming around, moving from one micro-colony to another.

A better understanding of biofilms is of high medical importance. Infections mediated by biofilms are tough to eradicate because biofilm inhabitants are up to 1,000 times more resistant to antibiotics than are free-floating bacteria. They are also less vulnerable to the immune system, and their matrix polysaccharides resist enzyme attack. Researchers initially thought the biofilm matrix provided a barrier to the diffusion of antibiotics. But in 1994, Gill Geesey and

his colleagues at the Center for Biofilm Engineering showed that antibiotics can penetrate the depths of a 500-micrometre-thick biofilm within 90 seconds¹⁹.

The true explanation for antibiotic resistance may be physiological: the lower metabolic rates of sedentary cells might make them less susceptible to drugs; the membranes of biofilm bacteria might be better equipped to pump out antibiotics before they can cause damage; or biofilm-dwellers may produce fewer of the proteins that are targeted by conventional antibiotics. In addition, biofilm bacteria exchange DNA much more readily than do free-floating bacteria, which might accelerate the transfer of antibiotic-resistance genes²⁰.

Thanks to the newly acquired knowledge on biofilms, these challenges may soon be overcome. Some believe that interfering with quorum-sensing signals holds great clinical promise. Researchers led by Michael Givskov at the Technical University of Denmark and Niels Høiby of the University of Copenhagen, for example, have infected rats' lungs with mutants of *P. aeruginosa* defective in two quorum-sensing systems. Clearing the microbes faster, the rats suffered much less tissue damage than controls²¹.

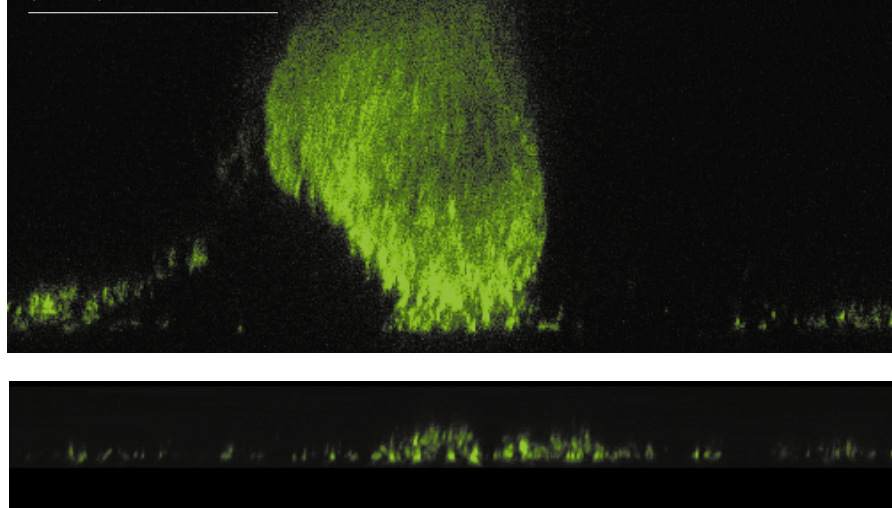
So the search is on for drugs that bind but fail to activate the receptors of quorum-sensing signals, molecules that interfere with signal synthesis, and enzymes that degrade the signals¹². "If you could trick bacteria so they couldn't count each other, and therefore they never knew when they had reached a high cell number, they wouldn't turn on all these virulence factors," says Princeton's Bassler. And as such drugs would target microbial pathways, they would be less likely to cause harmful side effects. In addition, because they would not kill bacteria, the selective pressure to develop resistance mechanisms should be weaker than with traditional antibiotics.

Slime management

Applying this approach to target a wide variety of bacteria, Quorex Pharmaceuticals in Carlsbad, California, is designing small molecules to inhibit the broad-spectrum signalling pathway discovered by Bassler. Screening their molecules' abilities to interfere with the production of virulence factors and biofilms, the researchers have identified several promising candidates to move into animal experiments.

Inspired by a natural inhibitor of quorum-sensing signals, researchers at the University of New South Wales in Sydney have also developed promising, broad-spectrum inhibitors. In the early 1990s, marine biologist Peter Steinberg was intrigued by the ability of the seaweed *Delisea pulchra* to remain slime-free despite living in bacterium-infested waters. Isolating compounds coating the seaweed's surface, he traced the antifouling activity to molecules called furanones. When

Film-struck: the mushroom structures of a *Pseudomonas aeruginosa* biofilm (top) do not form if bacterial signalling is blocked (bottom).



Steinberg's microbiologist colleague Staffan Kjelleberg saw the furanones' chemical structures, he had a flash of recognition: "They were basically exactly the same thing as a bacterial signalling system — acyl homoserine lactones."

Kjelleberg found that the furanones displaced the bacterial quorum-sensing signals from their receptors²². Working with researchers at the Centre for Marine Biofouling and Bio-Innovation and the company Biosignal, also in Sydney, Kjelleberg has created almost 100 synthetic variants of the seaweed furanones, hoping they will solve a wide range of biofilm problems. Initially, the researchers are developing coatings for keeping aquaculture equipment biofilm-free. Working with Givskov, Kjelleberg is also testing the furanones' effects in mouse models of human disease. "All I can say is that it looks promising," says Givskov. Again, drug resistance is not likely to be a problem. "*Delisea* has been using a signal blocker in nature for millions of years and no resistant bacteria have developed," says Costerton.

Focusing on more specific targets, Greenberg's team recently licensed a method for screening potential inhibitors of the quorum-sensing signals of *P. aeruginosa* to Quorum Sciences in Iowa City — now part of Aurora Biosciences in San Diego, California. The method is based on monitoring the expression of a subset of genes regulated by the signals. Greenberg is also developing drug-screening methods based on measuring ratios of quorum-sensing molecules. In a paper published in *Nature* last month, his team reported that the ratio of signals produced by *P. aeruginosa* living in biofilms differs from that produced by free-floaters²³. Indeed, by measuring these ratios in sputum, the team bolstered the hypothesis that many

of the *P. aeruginosa* that colonize the lungs of cystic fibrosis patients inhabit biofilms.

One key to developing further strategies for combating the formation of such biofilms may be to do as Costerton does, and try to consider these microbial cities from a bacterial perspective. "Microbes tend to be really smart," says Peter Hecht, chief executive officer of Microbia, a company in Cambridge, Massachusetts, that holds commercial rights to Kolter's work. "I guess that's why we want to think like them."

Marina Chicurel is a writer in Santa Cruz, California.

1. Costerton, J. W., Geesey, G. G. & Cheng, K.-J. *Sci. Am.* **238**, 86–95 (1978).
2. O'Toole, G., Kaplan, H. B. & Kolter, R. *Annu. Rev. Microbiol.* **54**, 49–79 (2000).
3. Costerton, J. W., Stewart, P. S. & Greenberg, E. P. *Science* **284**, 1318–1322 (1999).
4. Potera, C. *Science* **283**, 1837–1839 (1999).
5. Danese, P. N., Pratt, L. A., Dove, S. L. & Kolter, R. *Mol. Microbiol.* **37**, 424–432 (2000).
6. Pratt, L. A. & Kolter, R. *Mol. Microbiol.* **30**, 285–293 (1998).
7. O'Toole, G. A. & Kolter, R. *Mol. Microbiol.* **30**, 295–304 (1998).
8. Merz, A. J., So, M. & Sheetz, M. P. *Nature* **407**, 98–102 (2000).
9. Tolker-Nielsen, T. *et al. J. Bacteriol.* (in the press).
10. Prigent-Combaret, C. *et al. J. Bacteriol.* **181**, 5993–6002 (1999).
11. Parsek, M. R. & Greenberg, E. P. *Proc. Natl Acad. Sci. USA* **97**, 8789–8793 (2000).
12. de Kievit, T. R. & Igilewski, B. H. *Infect. Immun.* **68**, 4839–4849 (2000).
13. Davies, D. G. *et al. Science* **280**, 295–298 (1998).
14. Whiteley, M., Lee, K. M. & Greenberg, E. P. *Proc. Natl Acad. Sci. USA* **96**, 13904–13909 (1999).
15. Surette, M. G., Miller, M. B. & Bassler, B. L. *Proc. Natl Acad. Sci. USA* **96**, 1639–1644 (1999).
16. Nielsen, A. T., Tolker-Nielsen, T., Barken, K. B. & Molin, S. *Environ. Microbiol.* **2**, 59–68 (2000).
17. Picioreanu, C., van Loosdrecht, M. C. M. & Heijnen, J. J. *Biotech. Bioeng.* (in the press).
18. Xu, K. D., Stewart, P. S., Xia, F., Huang, C. T. & McFeters, G. A. *Appl. Environ. Microbiol.* **64**, 4035–4039 (1998).
19. Suci, P. A., Mittelman, M. W., Yu, F. P. & Geesey, G. G. *Antimicrob. Agents Chemother.* **38**, 2125–2133 (1994).
20. Hausner, M. & Wuerz, S. *Appl. Environ. Microbiol.* **65**, 3710–3713 (1999).
21. Wu, H. *et al. Microbiol.* (in the press).
22. Manefield, M. *et al. Microbiol.* **145**, 283–291 (1999).
23. Singh, P. K. *et al. Nature* **407**, 762–764 (2000).

Why are AIDS dissidents still making 15-year-old, long-refuted claims?

Sir—The attempt by Stewart *et al.* (*Nature* **407**, 286; 2000) to minimize the significance of the recent Durban Declaration (*Nature* **406**, 15; 2000) affirming that HIV is the cause of AIDS raises several troubling issues.

Stewart *et al.* suggest the bulk of the declaration's 5,000 academic signatories may have inadequate credentials. Yet the signatories of the Stewart *et al.* Correspondence, who make up most of the 'HIV-denialist' membership of President Mbeki's AIDS advisory panel, are mainly known for their disagreement on AIDS with just about everyone else in academic science, medicine and public health. In contrast, the signatories to the Durban Declaration include the vast majority of scientists worldwide who publish on all aspects of HIV and AIDS.

Second, Stewart *et al.* distort the declaration's statement that there is "no end in sight" to the epidemic by taking the phrase utterly out of context. Enormous progress has, of course, been made against AIDS, all of it stemming from the fundamental knowledge that AIDS is directly caused by HIV infection. This knowledge has enabled a long string of consistent and fruitful observations about the pathogenesis of HIV and brought about a powerful new family of pharmaceuticals that, however imperfect, have dramatically reversed the death rate from AIDS wherever they have been used.

There is "no end in sight" only where poverty, greed, politics or misguided information blocks access to these advances. Sadly, people in developing nations are paying the price with their lives, nowhere more so than in South Africa. Women are being deprived of safe, proven methods of blocking neonatal transmission, efforts to improve access to the new drugs are being undermined and years of prevention work are being confounded.

Third, Stewart *et al.* cite four papers from the early 1980s as a basis for objecting to the Durban Declaration. Whatever such papers say, they reflect only the knowledge available when they were written. Subsequent data have greatly refined our understanding. Why do Stewart *et al.* ignore 15 years of scientific progress?

Fourth, Stewart *et al.*'s claim that AIDS did not spread initially in Africa is simply incorrect. While the spread of HIV-1 and AIDS to sub-Saharan Africa was a later phenomenon, AIDS appeared in significant numbers more or less concur-

rently in several other African nations, Europe and North America in the early 1980s. Heterosexual transmission was evident almost from the beginning in Africa, as well as among transfusion recipients and haemophiliacs. Fear of an outbreak among Western heterosexuals was a valid concern. We were extremely fortunate in the United States to have a slow initial spread to heterosexuals, probably because the epidemic first broke out exponentially here among homosexual men, who do not routinely have sex with women. This observed pattern of HIV-1 spread is exactly what would be expected of a sexually transmitted disease with a first foothold in the gay male community. Today, heterosexual transmission is routine almost anywhere HIV-1 appears.

Although there may be some disparities in the ways AIDS affects people in developing nations compared with the West, there are no great mysteries. Higher rates of breast-feeding, for example, together with malnutrition and poor prenatal and delivery care, undoubtedly contribute to a higher rate of infant infection. The fact that many people in Africa have been unable to use prophylactic antivirals has also contributed enormously to the disparity in the rate of perinatal transmission compared with the West.

Finally, Stewart *et al.* argue that the HIV-denialists have had their views suppressed. On the contrary, scientists worldwide have shown excessive patience for the past 15 years. The accepted standard of science is to permit everyone to express their views, but also to hold people intellectually responsible for what they say. When a position is found faulty by the consensus of scientific opinion, principled dissenters go back to the lab and run new experiments in hopes of proving their point on a new day, rather than attacking the character of their critics and arguing their case to scientifically unqualified media and the lay public. The HIV-denialists are preaching to the very people at greatest risk now, the HIV-positive patient population itself.

Thousands of babies are born with HIV infection in South Africa alone, with little or no hope of a normal life. How high will the death toll have to be before the denialists see the error of their ways?

Martin Delaney

Project Inform, 205 13th Street, Suite 2001, San Francisco, California 94103, USA

Other signatories to this letter:

Linda Grinberg Foundation for AIDS and Immune Research, Los Angeles, California, USA

Mark Harrington Treatment Action Group, New York, USA

Lynn Morris National Institute of Virology, Johannesburg, South Africa

Mark A. Wainberg McGill University AIDS Centre, Montreal, Canada

John P. Moore Joan and Sanford I. Weill Medical College of Cornell University, Ithaca, New York, USA

Mildest organochlorines still cause toxic pollution

Sir—In his Correspondence about a recent book review, Ferdinand Engelbeen¹ criticizes Joe Thornton² for failing to distinguish between good and bad organochlorines, without proposing how one might do so at reasonable cost. One of the strengths of Thornton's book³ lies in its account of the exceptional difficulty of this process and the many ways we have failed to understand the implications of synthesizing chemicals that have no part in natural cycles.

Rather than undermining the case for restrictions based on chemical classes, Engelbeen's own example of PVC illustrates why it is so hard to regulate chemical compounds on an individual basis. Thornton argues that organochlorines, as a result of their fundamental chemistry, tend as a class to be greatly more persistent, more toxic and more bioaccumulative than their non-chlorinated relatives. Hence when degradation does occur it usually results in the production only of smaller organochlorines. Even organochlorines that are non-toxic or benign (such as PVC) are end-products of a sequence of reactions that at each stage result in the production of large quantities of toxic, persistent and uncharacterized chlorinated intermediates, by-products and wastes.

To make matters worse, even with the best available methods, organochlorines cannot be disposed of without the production of yet more toxins. PVC, via incineration, is probably ultimately the major environmental source of dioxins⁴. Nor is PVC always likely to be disposed of by the best methods. Significant quantities (even in the developed world) are disposed of in bonfires, building fires and rubbish-tip fires, and these too are likely to be important sources of dioxins and other chlorinated toxins⁴.

Jonathan R. Latham

6 Canon Frome Court, Canon Frome, Ledbury HR8 2TD, UK

1. Engelbeen, F. *Nature* **407**, 445 (2000).

2. Collins, C. *Nature* **406**, 17–18 (2000).

3. Thornton, J. *Pandora's Poison: On Chlorine, Health and a New Environmental Strategy* (MIT, Cambridge, MA, 2000).

4. Thomas, V. & Spiro, C. *Toxicology and Environmental Chemistry* **50**, 1–7 (1995).

Ancestors knew how to harness horsepower ...

Sir—Vaclav Smil's Millennium Essay "Horse power" (*Nature* **405**, 125; 2000) came as a pleasant surprise to me, as I would not have imagined that the readers of *Nature* were interested in this topic.

However, it is not true that “ancient throat-and-girth harnesses choked the animal”. This mistake was first made in the 1920s by Richard Lefebvre des Noëttes. His theory that slavery stemmed from the inability of the ancients to use animal power efficiently was very popular in his day. It lingers on in popular literature, although it has long been disproved: as Jean Spruytte showed in the 1970s, there were several ways of harnessing horses in antiquity, none of them choking the animals, and there is no connection between animal harnessing and slavery. (See, for example, Spruytte's *Early Harness Systems*, J. A. Allen, London, 1982.) Lefebvre des Noëttes had two different harnessing techniques mixed together in his mind. So Smil is right that the horse-collar was an improvement, but only an improvement on a existing technique.

Second, as far as one can tell, ancient and medieval horses were very small, often barely the size of present-day ponies. There is no incontrovertible evidence of the breeding of “heavy war animals needed to carry armoured knights” that has so often been supposed. In medieval times, large horses were a rare luxury. We lack the data on horse size to know what happened before the eighteenth century, so we cannot know whether armoured knights did ride big horses.

My last point concerns ploughs. The replacement of wood by iron and steel obviously allowed many improvements in the general structure and design of ploughs. But the case of mould-boards is special. The idea that “iron mould-boards only crossed from China to Europe in the seventeenth century” is speculative. There is no evidence of metallic mould-boards coming from China to Europe in time to be used as models by European makers of ploughs. (Chinese mould-boards, incidentally, were made of an alloy, cast iron, rather than of plain iron.) In Europe, wooden mould-boards were simply covered by more and more iron sheets to protect against wear. In some regions, wooden mould-boards were made with a curve from late medieval times.

Finally, wooden mould-boards had their own advantages. In the Gâtinais, north of Orléans, for example, arable soils are quite clayey and stick to iron mould-boards, whereas wooden mould-boards get soaked on their surface, forming a lubricating film of water that prevents the earth from sticking to it. Hence wooden mould-boards were used in this area even when ploughs were made completely of iron, up until the era of the tractor.

François Sigaut

Centre d'Histoire des Techniques, Conservatoire National des Arts et Métiers, 5 rue du Vertbois, 75003 Paris, France

... so animals could pull their weight, and more

Sir — Vaclav Smil's Millennium Essay “Horse power” presents a good overview of the important role of draught horses in agricultural production in North America during the last century¹. Some points worth adding are that mules were another key source of draught power on many farms², and that the larger draught-horse breeds are, during brief exertions, capable of developing even more than the three horsepower Smil mentions.

Records of draught-horse championship pulling trials in the United States show that a team of two animals could develop 30 horsepower when pulling loads over a set distance³. Similar performances have been recorded for teams in Europe and elsewhere. Average working performance for one horse is 0.75 to 1.0 horsepower⁴.

Michael R. Goe

Department of Animal Sciences, Swiss Federal Institute of Technology, Tannesstrasse 1, ETH-Zentrum, 8092 Zurich, Switzerland

1. Smil, V. *Nature* **405**, 125 (2000).
2. Leaflet 225 “Quick Reference Map of Horses and Mules by States” (Horse and Mule Association of America, Chicago, 1936).
3. Collins, E. V. & Caine, A. B. *Testing Draft Horses* (Iowa Experiment Station Bulletin 240, Ames, 1926).
4. Goe, M. R. & McDowell, R. E. *Animal Traction: Guidelines for Utilization* (Cornell International Agriculture Mimeograph 81, Ithaca, NY, 1980).

Fraud: retracted articles are still being cited

Sir — Much heat continues to be generated by the topic of scientific fraud. In the United States, the Office for Research Integrity identifies papers in the biomedical sciences declared to be fraudulent after an official inquiry (<http://ori.dhhs.gov>) and publishes their bibliographic data.

In addition, papers are often retracted from journals after publication for other reasons. But are these data being generally disseminated?

There are some grounds for pessimism. Kochan and Budd¹, for example, showed that retracted papers by John Darsee continued to be positively cited even though a considerable amount of time had passed since retraction, and even though the case generated much publicity. Pfeifer and Snodgrass² recorded citations to 82 completely retracted articles and found that, although retraction reduces subsequent citation compared with a control group, retracted papers were often cited to support claims. Finally, Budd *et al.*³, using Medline to identify articles retracted between 1966 and 1997, found

that many retracted articles were still being cited as valid.

It seems, therefore, that in at least some cases, authors are not aware of retractions.

A systematic screening method is required to prevent the citation of fraudulent or retracted papers. This could be done for some disciplines via databases such as Medline (which can be searched for retracted publications), but would not cover all fields of research.

Another approach would be for an organization, for example a scientific publishers' group, to run a web-based database of retracted and/or fraudulent papers covering all fields of research. Authors could then search this database before submitting their papers for publication. This search could be a requirement for submission of the paper to a journal.

Such a database could also help people to avoid doing new research based on useless claims in the literature.

How such a venture would be funded, and how it would work in practice, are the next questions to address.

Juan Miguel Campanario

Departamento de Física, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain

1. Kochan, C. A. & Budd, J. M. *J. Am. Soc. Inform. Sci.* **43**, 488–493 (1992).
2. Pfeifer, M. P. & Snodgrass, G. L. *J. Am. Med. Assoc.* **263**, 1420–1423 (1990).
3. Budd, J. M., Sievert, M. E. & Schultz, T. R. *J. Am. Med. Assoc.* **280**, 296–297 (1998).

The Editor replies — *Nature* requires authors to have performed appropriate checks before submission to ensure the validity of references cited as far as is possible. *Nature* also has a policy of publishing retractions and corrections. These are linked to and from the relevant paper on the *Nature* website.

Award organizers should have noted the paper

Sir — In the News story “Ig Nobel glory for levitating frogs and collapsing toilets” (*Nature* **407**, 665; 2000), my name, listed as one of the co-winners of the Ig Nobel psychology prize for the paper “Unskilled and unaware of it: How difficulties in recognizing one's own incompetence leads to inflated self-assessments”, was misspelt — as it is on the Ig Nobel website. It is Kruger, not Kreuger.

This accident is hardly surprising, however, in light of our topic. It can be very difficult to spot one's own mistakes.

Justin Kruger

Department of Psychology, University of Illinois in Urbana-Champaign, 603 East Daniel Street, Champaign, Illinois 61820, USA

Technology's tortoise and hare

The sociological dynamics are now right for the electric car to eclipse its rival.

The Electric Vehicle and the Burden of History

by David A. Kirsch

Rutgers University Press: 2000. 256 pp. \$20

Stanford Ovshinsky

History is not a linear progression of events. The history of science and technology is packed with unexpected basic advances that are not recognized until they have reached a critical mass and then surged forward to change the way we live, work and understand the world around us.

We live in an age when the foundations of the global economy — energy, transport and information systems — are being transformed. In the United States Jack Smith, chairman of General Motors, and Bill Ford, chairman of the Ford Motor Company, agree that no car manufacturer will go far in the new century with the internal-combustion engine (ICE). It is hardly necessary to survey the serious societal problems underlying these opinions. They are emphasized by the current global oil crisis, with its profound consequences for the world economy.

The transport system based on the ICE is undergoing a transition to the use of electric, hybrid and fuel-cell vehicles. The dependence of these vehicles on hydrogen — in metal-hydride batteries for electric and hybrid vehicles, or as the fuel for fuel cells — is initiating a new age: the age of the hydrogen economy.

How can we understand the emerging electric-vehicle industry? The simplest approach might appear to be to study the events of 100 years ago, when electric vehicles were overtaken by the incredible growth of the automotive industry, as the success of the ICE with an electric starter-motor made cars practical and affordable.

However, what if such a history turned out not to be truly relevant to what is unfolding now? David Kirsch's book is an informative history, full of excellent case studies of the various attempts to build a transport system based on battery-driven vehicles. Its strength lies in the fact that it takes a systems approach, combining social, environmental and business perspectives to provide a well-researched analysis of the early battle between the ICE and battery-powered vehicles. As history, it is an excellent, insightful book.

However, as the book reaches modern times, it lacks both technological depth and political understanding. Kirsch does not appear to be up to date with California's struggle to enforce the mandate set by its Air Resources Board that, by 2003, 10% of

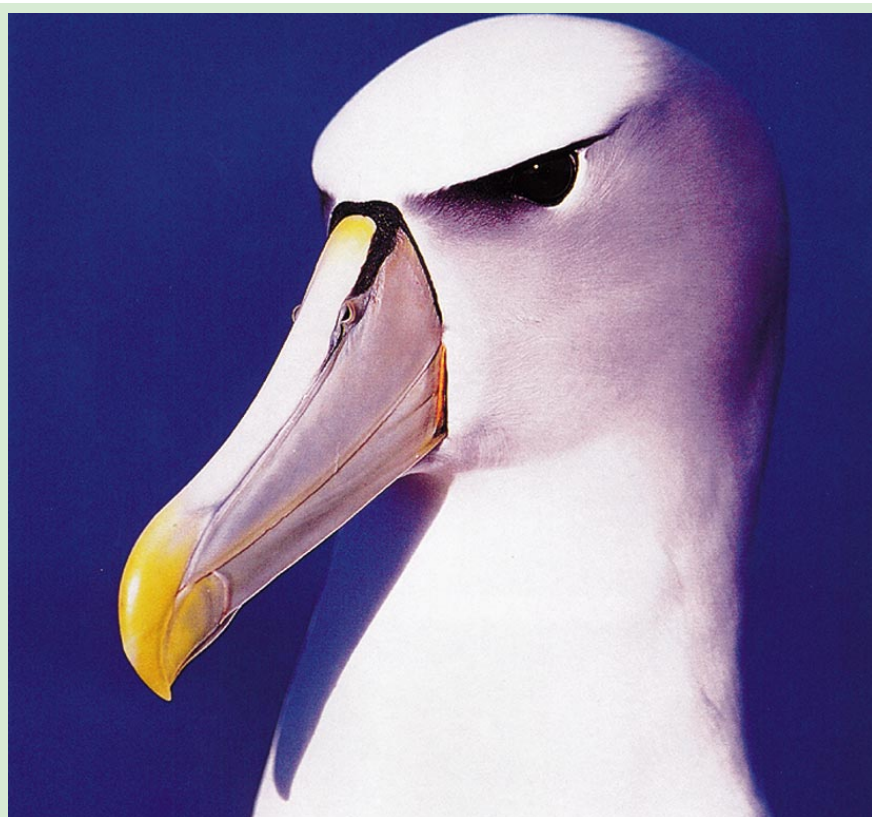
vehicles sold in the state should produce no emissions — despite heated opposition from various groups. The mandate will be enforced by the Air Resources Board from 2003. It is expected that car companies will comply rather than face possibly severe monetary penalties.

The problem now is not, as Kirsch believes, whether a breakthrough in battery technology has provided a dramatic increase in range, nor whether there is a market for electric vehicles, nor whether costs can compete with those of internal-combustion vehicles. The crucial hearings of the California Air Resources Board this September answered all of these in the affirmative so definitively as to result in a unanimous vote to continue the mandate.

Past is not prelude. Just as many generals fight the previous war, so it is a mistake to think that the present situation reflects the sociological dynamics of the turn of the last century. Heraclitus was right: it is not possible to step twice into the same river.

What is involved is the task of changing a huge, powerful, entrenched global industry, with enormous resources as well as financial and psychological investments, from a petrol-engine base to an electric one. It is a dynamic process of competing strategies and, for the automotive industry, a very traumatic event in a vital historical drama. The industry's attitude can be likened to that of St Augustine: "Make me chaste, Lord, but not yet."

A recent report from the US Department of Energy provides data showing that



A burden undeserved

Far from being the encumbrance and sign of retribution portrayed in Samuel Coleridge's *Rime of the Ancient Mariner*, an albatross is a welcome and impressive sight — and one that Coleridge himself never experienced. The so-called shy albatross shown above belongs to this diverse grouping of seabirds, which inhabits the Antarctic and Pacific oceans. *Albatrosses* by W. L. N. Tickell (Yale University Press, \$60), from

which the picture is taken, is a comprehensive account of the 13 groups of albatrosses, and compares their natural history, breeding, flight and ecology. Measures to reduce the numbers that in the past have been inadvertent victims of the fishing industry are also described. The book's final chapter looks at some poems in which these birds have figured; it is called "The Mariners syndrome".

electric vehicles propelled by nickel metal-hydride batteries based on new scientific principles can achieve a range of well over 200 miles, answering the question of viability of pure electric vehicles and hybrids. The electric vehicle's superiority in city driving was demonstrated in 1998 during a mileage comparison test between two Geo Metro vehicles, one converted to run on nickel metal-hydride batteries, the other a conventional petrol version. In heavy New York city driving conditions, the electric version demonstrated a projected range of 362 kilometres, as against 192 kilometres for its ICE counterpart.

The Institute of Electrical and Electronics Engineers monitored a 347-kilometre trip from Boston to New York city by a four-passenger Solectria using similar batteries. It completed the journey on a single charge, with 15% energy remaining — having used the energy equivalent of less than one gallon of petrol.

Impressive as these results are, they are not the whole story. For, as the recurring problems of availability and affordability of petrol have shown, a city must still function. Global pollution and climate change have made the electric vehicle a necessity as a delivery vehicle, taxi or bus. In addition, hybrids, which can do 100 kilometres on 3.5 litres or less (80 miles or more to the gallon), using a small petrol-based engine and similar nickel metal hydride batteries, are becoming part of the accepted vehicle mix so necessary in establishing an automotive industry. Even now, the ICE automobile needs greater battery power to fuel the increasing number of electrical components it contains.

The question of affordability has been answered by Robert Stempel, former head of General Motors, the world's largest corporation. Stempel, a pioneer whom many consider to be the best automotive engineer in the industry, is the father of the first commercial electric vehicle, EV1. He gave testimony to the California Air Resources Board that an electric vehicle made in the same numbers as conventional cars today would be cheaper to produce than its petrol-based counterpart. Until such volumes are achieved, the industry should use the time-honoured mechanism of 'forward pricing' — introducing a new product at an affordable price so that it can achieve sufficient volume to become profitable.

It is clear that events have overtaken academic history in this field. The new electric vehicle does not carry the burden of history. On the contrary, in its various manifestations it is spurring new scientific, technological and social inventions and is the vehicle for the kind of change that is desperately needed in the new millennium. ■

Stanford Ovshinsky is at Energy Conversion Devices, 1675 West Maple Road, Troy, Michigan 48064, USA.



Financial pressures: econophysicists have not yet eased the unpredictability of the stock market.

Physics of the money markets

An Introduction to Econophysics

Rosario N. Mantegna & H. Eugene Stanley
Cambridge University Press: 1999. 158 pp.
£25, \$34.95

Blake LeBaron

The arrival of high-quality machine-readable data sets nearly 40 years ago transformed finance into the most quantitative social science. Since then, finance has seen a steady influx of talented researchers from the natural sciences and applied mathematics. The field of econophysics marks a key distinction from these previous incursions, as it is trying to establish an entirely new field, separate from that of financial economics.

The book by Rosario Mantegna and Eugene Stanley is an attempt to define this field by two of its earliest proponents. It is also an introductory tutorial for physical scientists who are just becoming interested in finance. In this second role, it is a very useful overview.

Its more formidable problem is in defining the bounds of econophysics as a field, and introducing its tools. The book reflects the authors' research on new empirical techniques. Using recently available high-frequency financial data — data obtained by monitoring changes in share prices at very short intervals — the authors put series of minute-by-minute share-price movements for several companies under a physics-oriented microscope. Some of these series contain nearly half a million data points sampled at one-minute intervals.



The authors show that many series of financial share-price movements are close to a 'random walk' — in agreement with traditional financial wisdom that markets are difficult to forecast. They also confirm that, when sampled at short intervals, the increments to the random walk are irregular, with sharp peaks and fat tails. This means that there are too many large and small changes in share price for the movements to be coming from a normal distribution. Finally, the authors show that the magnitude of these price changes is very persistent, that is, that large changes in share price, either up or down, are followed by further large changes. These three features are well known for share prices monitored at longer time intervals, and can be traced back to the 1960s work of Eugene Fama and Benoit Mandelbrot.

What distinguishes this empirical work from traditional finance is its emphasis on scaling relationships in the data. The authors' scaling results show that a comparison of high- and lower-frequency financial data reveals some interesting similarities. In

other words, a researcher analysing the properties of a series would not be able to distinguish data recorded at one-minute intervals from those recorded every 30 minutes (adjusted by an appropriate scale factor). If this were true, it could have far-reaching implications for economic theory in that it suggests a kind of continuum in behaviour from short to long timescales.

Although interesting, the results are difficult to interpret. First, it is not clear what types of processes might generate pictures that would have similar properties to the actual data. The book gives some examples, but it is not easy to see how visual scaling laws alone can distinguish between the different candidate processes for representing stock returns. Second, the statistical properties of the scaling relations must be shown to be universal and reliable. Statistical tools such as monte-carlo and bootstrapping can give useful information in such situations. Also, simple tests that could demonstrate the stability of the scaling parameters over various time periods, and for many companies, could offer a convincing case for scaling universality.

The scaling does break down eventually. On longer timescales (roughly 30 days), price changes approach a normal distribution. This feature is well known in finance, but is disturbing for econophysics. The idea of using common distributional features to define theories at all timescales is not in line with the empirical facts, and this weakens the usefulness of scaling relationships as a guide to economic theorizing.

In terms of coverage, I was disturbed by two omissions. First, a field called econophysics should cover more than finance. Work has been done on other time series — such as company growth rates or individual income distributions — by the authors and others, and it would have been useful to see it compared with the financial results. A unification of scaling properties across diverse socio-economic time series would be a very powerful result. Also, the field of econophysics contains work on agent-based, or microscopic, markets, which looks at how individual market traders interact. These models are the theoretical counter for some of the observational results in the book. This work is cited, but is given very little coverage. Such theoretical models are radically different from traditional finance theory, and definitely help to show where econophysics differs.

In summary, I feel the book is a useful introduction to the empirical aspects of econophysics, and it will be well received by natural scientists interested in an introduction to finance, but those already familiar with financial time series will not find this a ground-breaking read. ■

Blake LeBaron is at the Graduate School of Economics and Finance, Brandeis University, 415 South Street, Waltham, Massachusetts 02453-2728, USA.

A cosmological bouquet

The Book of the Cosmos: Imagining the Universe from Heraclitus to Hawking

edited by Dennis Richard Danielson
Perseus Publishing: 2000. 556 pp.

\$35, £21.50

Giovanni F. Bignami

Students are pleasantly surprised when they learn the etymology of 'anthology' — picking flowers from the meadows of a given subject. And this anthology on cosmology begins by noting another etymological kinship — that between 'cosmic' and 'cosmetic': both are related to order and beauty. So, what better, more vast and more challenging meadow than cosmology from which to pick your flowers, especially for someone like Dennis Danielson, who is both fortunate and bold enough to teach courses in the 'literature of cosmology'?

But what should the anthology consist of? Should the anthologist represent what grows in the meadow, his own tastes, or should he try to cater for those of the reader? Covering a vast range in time and tastes, as Danielson does here, seems a safe principle. The result is 85 chapters, each, according to Danielson, being a "telescope for the mind".

The ancient Greeks are solidly and appropriately represented in the book, and some effort has gone into rendering their thoughts accessible to literature-of-cosmology students and general readers alike. The problem is, of course, that the Greeks, like Cicero in Shakespeare's *Julius Caesar*, spoke Greek. In fact, this book is very much about translations into English, because most of mankind's contributions to cosmology were not originally written in English. The many translations in *The Book of the Cosmos* are clear and precise, and, impressively, some are by Danielson himself. Some are even of significant literary value, as in the contribution by the English translator Thomas Creech — "*Lucretius ... De Rerum Natura Done into English Verse*" (1683); a true gem.



Arab eye: the Muslim philosopher Averroës had ideas on cosmology that clashed with those of his culture.

ARCHIVO ICONOGRAFICO, SA CORBIS

In the context of translations, Danielson has missed a good opportunity here to correct the long-standing error in the translation of Galileo's *Sidereus Nuncius* as 'Starry Messenger'. Galileo himself wrote very clearly that *Nuncius* was intended to be as 'message', not 'messenger'.

The past is, by definition, politically correct. Arabs existed in the past (as they do today), and Arab thought, especially in astronomy and cosmology, is very much part of Western culture, as the names of many stars remind us every night. The Muslim philosopher Averroës, to quote but one, had strong ideas about cosmology. And because of them he ran into trouble with his own religious authorities, which went by the literal interpretation of the cosmological aspects of the Koran (also entirely neglected in Danielson's book). Considering the initial emphasis the book places on the Torah of the Jews, the absence of Arab authors and thought in this anthology gives an incomplete picture of our direct cultural inheritance.

Danielson does include some irritating digressions in his descriptions of the classics. Why, for example, do we read so much about Arthur Koestler in a chapter supposedly about the Greek philosopher Parmenides? And why should Galileo's translated words be interspersed with rather irrelevant quotes by the nineteenth-century astronomer Mary Agnes Clerke? They are far less acceptable than Danielson's own comments, which are not only required by the genre, but also quite useful.

The framework containing Nicolaus Copernicus, Giordano Bruno, John Milton, René Descartes, Isaac Newton and Immanuel Kant, and also Mary Agnes Clerke and all the others, is very useful. So, in general, is the choice of morsels presented.

Inevitably, though, things become more complex with contemporary twentieth-century cosmology. Opposing schools of diverse thought thrive in our global village, no longer truly 'Western' as the United States has come to embody global cultures. English, like Greek, Arabic, Latin and French before it, is now *de rigueur*, and this simplifies things. Deceptively so, however, because modern cosmology retains uncertainties and debates (origin of redshifts), and a diversity of voices, including non-Anglo-Saxon ones (where is *tavarisc* Zeldovich?) that don't make it onto the pages of this book. Well, as the Romans used to say, *tot capita, tot sententiae* [loosely, each to his own, especially his own taste]. As for the definition of 'book'. This is given in the glossary and is obviously much needed, as it can range from a "device which takes the form of bound pages" to "the universe", as in "the book of the cosmos". ■

Giovanni F. Bignami is at the Italian Space Agency, viale Liegi, 26 00198 Roma, Italy.

Science in culture

A window on the future

The 'Museum for the Third Millennium' has opened in Melbourne, Australia.

Mark A. Elgar

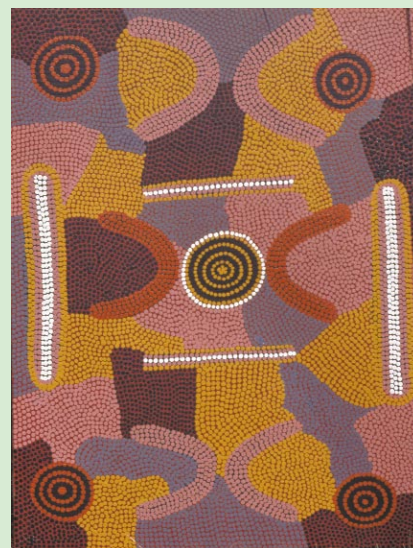
The largest museum complex in the Southern Hemisphere opened last month in Melbourne, Victoria's capital city. Melbourne Museum is the flagship campus of Museum Victoria, which is responsible for the state's extensive collections. Located for the past 100 years in the cramped State Library Building, the new museum complex is adjacent to the historic Royal Exhibition Building in the picturesque Carlton Gardens.

The relocation to a new, larger building also coincides with a substantial shift in thinking about the arrangement and design of museum exhibits. The shift has been from passive displays of historical artefacts and stuffed animals to exhibits of particular themes that place the collections in a broader context. According to chief executive George MacDonald, the museum will be "full of life and activity, a museum for the Internet Generation and beyond. A museum for the future, not the past."

And they are bringing life to the museum. The Forest Gallery is a world first; a living interpretation of Victoria's tall mountain forest environment, housing 8,000 plants of at least 120 species, and home to a variety of animals, including birds, reptiles, small mammals, frogs and numerous invertebrates. These living 'exhibits', together with various multimedia installations, tell stories about how past and present forces of change have shaped the forest habitat. Unlike zoo exhibits, which focus on a single species, the Forest Gallery provides a more holistic view of the environment. Ross Field, environment programme director, believes the exhibit will "encourage a deeper understanding and appreciation of Victoria's temperate forests, and stimulate visitors to go out and visit them".

The museum delivers messages around issues, rather than the traditional focus on taxonomic groups or historical eras. The exhibitions relate to six core programmes that are the current focus of research and collection development (science, technology, Australian society, indigenous cultures, human mind and body, and environment). Thus, the Australia Gallery shows how Melbourne has evolved, and includes the mounted hide of an Australian sporting icon, the legendary racehorse Phar Lap. Bunjilaka, the Aboriginal Centre, provides a conduit for the Aboriginal communities of southeastern Australia to tell their stories, and allows visitors to explore the nature of Aboriginal law, property, knowledge and their complex relationship with land.

The change in museum philosophy is more than just rearranging exhibits. Robin Hirst, director of programmes, research and collections, says that, "traditionally, museums hold up a



Painting place: Aboriginal art by Paddy Karol.

mirror to reflect the past, while our new museum holds up a window through which we can look forward to what the future might hold". This philosophy is evident in the astonishing Body and Mind Gallery, which opens in a few months' time. This is no stamp collection in a dusty cabinet, but an impressive use of both collections and local expertise to address contemporary issues and provoke public debate. Visitors gain insights into the processes, structures and functions of the mind and body through some 1,300 objects, imagery and interactive experiences, including a series of full-size nude photographs and preserved sections of real cadavers. Undoubtedly, these exhibits will be controversial.

A museum is only as good as its collections, curators and research staff; otherwise, it is little more than an exhibition hall. The move to Carlton Gardens finally brings research staff into the same building as the exhibits. Previously hidden away in a shabby annex, they are now highly visible, through the plate-glass walls, in their well-appointed offices and laboratories.

And here's the rub. It cost some A\$290 million (US\$150 million) to develop the museum and its very impressive exhibits, of which \$263 million came from the Victoria state government. The museum is quick to acknowledge that the programmes, collections and research are an integral part of the exhibition development process. But although the new building provides enviable infrastructure, the museum offers little more than token funding for its research staff. Instead, researchers, like their university colleagues, must compete for government and private-sector funds, which are modest in Australia. A Museum for the Third Millennium may need to make more than a rhetorical investment in its future. ■

Mark A. Elgar is in the Department of Zoology, University of Melbourne, Victoria 3010, Australia.

Nice work — but is it science?

Untestable ecological theory won't help solve environmental problems.

Jim Smith

Scientists often dismiss philosophy as a vague, incomprehensible subject bearing little relation to science's 'real world' knowledge. Philosophers, on the other hand, have long debated the basis of scientific knowledge and how 'good' science is practised.

The most significant development in the philosophy of science in the past century was Karl Popper's book *Conjectures and Refutations* (1963), in which he showed how science can be distinguished from non-science or pseudo-science. Popper argued that all scientific theories are no more than conjectures that are more or less well tested. Any theory may be disproved by observation or experiment and replaced by a new, better theory.

The classic example of the working of good science is relativity theory's overthrow of newtonian mechanics. Relativity was first tested by its prediction that the Sun's gravitational pull would bend starlight, a phenomenon not predicted by Newton. The spectacular agreement between Einstein's predictions and observation of the 1919 solar eclipse were both a refutation of newtonian mechanics and the first test of relativity theory.

Popper emphasized two characteristics of scientific theory. First, a theory makes predictions that can be tested by empirical observation; and second, it is provisional: no matter how often it is tested, it is always possible that a future experiment or observation will contradict it. We cannot claim that relativity theory represents absolute truth because new observations may someday show it to be false. Despite this, we have faith in scientific theory because it can be tested against real data. Non- or pseudo-science, to Popper, is knowledge which is not, or cannot be, tested.

Issues concerning hypothesis testing are particularly relevant for the environmental and ecological sciences, because the complexity of environmental systems often limits our ability to test hypotheses quantitatively. Much 'theoretical' environmental and ecological science is necessarily based on laboratory experiments carried out under very controlled conditions. But little effort goes into extrapolating the results of these studies to the outside world. Theoretical ecologists studying fruitfly populations in the lab would be horrified at the thought of their carefully calibrated models being used to predict the behaviour of wild populations. The assumptions relevant to the laboratory would not hold in a real, complex environment. The models are therefore very weakly

Novelty and innovation are often rewarded at the expense of more useful applied research.

predictive. Of course there is a place for detailed studies of environmental processes, but such studies will not constitute useful scientific theory until they can be used to make predictions in the real world.

The most pressing environmental problems require scientists to predict the effects of man-made changes. We need, for example, to predict the density and distribution of artificially introduced species; to evaluate the impact on an ecosystem of a road-building programme; or to assess the consequences of pollution. Despite the urgent need for properly tested predictive models, such models are extremely rare. Although the scientific literature teems with mathematical models, these are almost never applied and tested by formal 'blind' predictions of the real-world behaviour of environmental systems.

In his *Critique for Ecology* (1991), US ecologist Robert Peters argued that environmental and ecological scien-

tists have focused on theoretical 'insight' and 'explanation' of complex ecosystems at the expense of predictive solutions to applied problems. Peters believed that "because many of the central issues of ecology are confused or untestable, ecologists are uncertain about what information the science needs". This has resulted in a 'trap of originality' in which novelty and innovation are often rewarded at the expense of applied research which, although it may be less technically brilliant, may produce results more useful to society. Ingenuity and innovation may be characteristics of good science, but they are not a substitute for properly elaborated and tested theory.

Faced with a complex environment, scientists have responded by developing ever more intricate and detailed techniques and models. Yet, ironically, simple semi-empirical relationships are usually a much better guide to the behaviour of environmental systems. Such models have achieved some notable successes: predicting a lake's biological response to phosphorus, for example, was key to justifying, and managing, phosphorus removal from effluents. Similarly, empirically based predictions of habitat requirements guided the successful reintroduction of the large blue butterfly (*Maculinea arion*) to Britain. Perhaps ecological and environmental scientists should concentrate less on theoretical explanation and more on finding applied solutions to humankind's environmental problems. ■

Jim Smith is at the Centre for Ecology and Hydrology, Winfrith Technology Centre, Dorchester, Dorset DT2 8ZD, UK.



Keep it simple: empirical studies of habitat requirements were the key to saving the large blue butterfly.

JEREMY THOMAS

At the zoo

The last two baseline humans in captivity have bred successfully.

Warren Ellis

28/03/2468 2225.18 Universal Time Code. The Human Reservation, in temperate London, England, considers the breeding of the last two baseline humans a vindication of its hard work in identifying, isolating and protecting the 'pure', second-millennium strain of human being.

"The cusp between second and third millennium is vital," explained the baseline humans' head keeper, Daybreak Sigridsdottir, in a brief radiotelepathy interview conducted during the act of disposing of the couple's physical waste products, known as 'mucking out'. "The emergence of practical biotechnology happened right at that point. The introduction of artificial and alien material into the human genetic structure began barely ten years into the new millennium. You can't trust anything after 2001."

Sigridsdottir is an accomplished zoologist, whose text about the rescue and subsequent massive sexing of the population of New York City after the release of the Genital Babel terrorist bioware won her the William Gates VI Truth Prize for 2461. Her work was crucial in sifting through world records to locate humans who remained utterly unchanged by modern biological technology. She well remembers the conflict of feelings when it was discovered that only two humans on Earth still possess a pre-2000 biological make-up.

"We stopped thinking about the way we live a long time ago. We take a deep breath to oxygenate enough for winged flight, but

we don't consider what we left in that lungful of air. In 66 cases out of a hundred, a cubic metre of sea-level air will contain an airborne genetic trait, engineered in a lab or spored from a working organ, capable of applying itself to the biology of the respiring individual.

"And now here we are. Most of us are at least 0.5% of our DNA away from the human baseline. Some are more. On the other side of the human species, if you go a percent or two of DNA away, you have a chimp. By the same token, many of us are no longer human. We are something else entirely. And the horror and surprise and sadness that has been expressed are artefacts of our refusal to accept that."

Nine Nevada Rockets, spokesgroup for the Transhuman Association, relays a different perspective on the event from its spawning territory in the Pacific Ocean. "This is nonsense. I could make a case for the genome being 'untrustworthy' post-1945, with the introduction of trace elements from nuclear detonations into the environment. Human genetic structure changed with each new immunity we developed and passed down to our descendents. Who is the Human Reservation to say what is human and what isn't?"

Sigridsdottir, who points out that Rockets is responding via a telepathy organ derived from whale and eel biologies, responds: "Look at us. We've adopted mechanical traits, we've taken on animal traits, we've invented new abilities and new internal organs to perform them from scratch. Rockets is nine neo-dolphin bio-

forms sharing a single brain on a local-access network run by organic modems located in their beaks. We are as far removed from these two people as they are from chimpanzees in the wild."

The unnamed couple, existing in a large and heavily secured enclosure dressed to resemble the urban environment they were found in, have not had an easy captivity. They have endured terrorist attempts to corrupt their genome, abortive assassination strikes, and even bomb threats.

One suspect in the latter is Laura Magdelene Manson, amanuensis for the non-denominational Assembly Of Mysterious Devotion, a barely visible cloud of femtotechnological machinery: "We were created to adapt and change. It was our appointed role to take on the traits of animals — to contain within ourselves all the abilities of the machines we have conceived. We do this to become worthy companions of The Mystery. To propagate and preserve inferior iterations of the human being is sick."

Sigridsdottir is dismissive. "It's all part of the ongoing psychosis of the culture. Despite the fact that we take for granted physical abilities that the second-millennium human could find only in *outré* fiction, we somehow feel threatened by the notion that someone is 'purer' than us. And the Assembly hasn't been the same since it was banned from using the word 'God' and other inflammatory antique terms. Preservation is a vital function of a civilized culture."

Why was physical, 'classic' breeding required? "Because it's now almost impossible to guarantee the absolute purity of a laboratory environment. And even if we could clone them in complete protection — that kind of defeats the point, to me. Using the classic method seems to me to be the most aesthetic way to accomplish this. It's the organic way. And, personally, I found it fascinating. No-one in my family has had actual, traditional, penetrative sex in seven generations. And now I see why."

So how difficult was it to implement the breeding? "Well, it wasn't as easy as you'd think. Sperm counts were remarkably low in the late-twentieth-century human male. The genetic imperative to reproduce was slowing over much of the world due to the rise in live births and changes in societal mores. Frankly, we had to get her drunk and promise him a pizza."

Warren Ellis (<http://www.warrenellis.com>) is the author of the award-winning Transmetropolitan series of graphic novels and various other works in the comics field.

JACEY



Plotting the pyramids

Owen Gingerich

There is no evidence in ancient texts that Egyptians used astronomical knowledge in building the pyramids. But analysis of the night sky in 2500 BC could help explain how the pyramid builders knew the direction of true north.

The Egyptian pyramids at Giza — one of the seven wonders of the ancient world — are roughly 4,500 years old. Their construction can be dated, at best, to within 100 years by existing methods. In this issue (*Nature* 408, 320–324; 2000), Egyptologist Kate Spence proposes a more accurate way to date these monuments using subtle deviations in the alignment of their bases from true north. At first glance, this idea seems as likely as sharpening razor blades by placing them under small pyramids (one of the claims made for ‘pyramid power’ in the 1970s). But on closer inspection, Spence has come up with an ingenious solution to a long-standing mystery: how were the great pyramids so accurately aligned in relation to north?

Since the nineteenth century it has been known that the western and eastern sides of the Khufu (Cheops) pyramid deviate from true north by an average of three arcminutes. This corresponds to one-tenth the diameter of the full moon and is not far from the 1–2 arcminute precision achieved several millennia later by Tycho Brahe, the greatest of the pre-telescopic observers.

Such an accurate alignment is possible only by some astronomical procedure, but the surviving Egyptian texts seem silent on the method used. One way to define north would be to compare the rising and setting positions of the Sun in the east and the west and bisect the angle between them. Such a procedure would have to be carried out near the time of a solstice, when the Sun seems temporarily to stand still in its seasonal march higher or lower in the sky. But problems with observing objects near the horizon (because of interference from the Earth’s atmosphere), and with having a perfectly level view both east and west, make this technique unreliable.

An alternative method would involve building some sort of scaffolding to provide a sight line for alignment with the north pole star — except there was no suitable pole star when the pyramids were built. Directly above their terrestrial counterparts, the celestial poles are points in the night sky around which the stars appear to rotate. Today, the north celestial pole points close to the star Polaris — currently designated the pole star. Because the Earth’s axis of rotation is not fixed, but swings in a stately conical motion known as precession, the positions of the celestial poles drift slowly among the

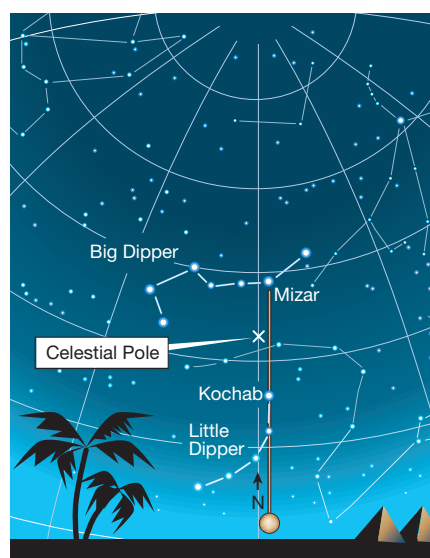


Figure 1 The night sky in ancient Egypt. Kate Spence argues on page 320 of this issue that Egyptian pyramid builders used the vertical alignment of two bright stars in the constellations of the Big and Little Dipper to align their pyramids north–south. In around 2500 BC, the star Mizar in the Big Dipper and Kochab in the Little Dipper were located on opposite sides of the north celestial pole — the fixed point in the sky around which all the stars rotate. In 2467 BC an invisible line linking these two stars, when they are vertically aligned, would pass precisely through the north pole. A plumb line intersecting these stars would then point straight to north on the horizon. A similar observation made before or after this date would include a systematic error caused by the alignment being offset slightly from true north (as is the case here). Such astronomical errors show up in the orientations of the ancient pyramids of Giza.

stars in a 26,000-year cycle.

This is where Spence’s idea comes in. Although there was no pole star available for accurately pinpointing north in ancient Egypt, there was a pair of fairly bright stars, one either side of the ancient celestial pole, which in 2467 BC lay precisely along a straight line including the pole. One is Kochab in the bowl of the Little Dipper, the other Mizar in the middle of the handle of the Big Dipper (to use the popular American names for Ursa Minor and the main stars of Ursa Major). Kochab and Mizar are technically known as β -Ursae Minoris and ζ -Ursae Majoris. In

2467 BC, an Egyptian astronomer could wait while the heavens slowly pivoted around the unmarked pole until a plumb line exactly intersected both stars, one above the invisible pole and the other below it (Fig. 1). The sight line to the horizon point directly below the plumb line would then point straight to north.

So far Spence’s suggestion is as speculative as it is ingenious. But she then introduces a second, hitherto unrecognized, mystery. The great pyramid of Khufu is the most accurately aligned to north, deviating just slightly to the west. The pyramid of Menkaure, the second pharaoh after Khufu, errs by about 13 arcminutes in the other direction, whereas the three earlier pyramids of Snofru, Khufu’s predecessor, are oriented more westward as they get older. Other pyramids, built between 2600 and 2300 BC, confirm this trend, and their orientations form a straight line when plotted against time.

How can this puzzle be explained? Because of the Earth’s precession, the celestial north pole was exactly aligned between Kochab and Mizar only in the year 2467 BC, and the errors in the orientations of earlier and later pyramids faithfully track the slow drift of Kochab and Mizar with respect to true north. Because the error in the Kochab–Mizar alignment can readily be calculated for any date, the error in each pyramid’s orientation corresponds to a specific year. Although each alignment is subject to individual measurement errors, the collection of several data points makes the method more robust. So it is not preposterous to believe that Spence can calculate dates for pyramid construction to within five years or so, considerably better than the 100-year error currently accepted for their chronologies.

A modern astronomer noticing the drift of the two stars could easily come up with a simple observational solution to the problem. After 2467 BC, with Kochab directly above Mizar, the alignment missed the pole to the left. But 12 hours later, with Mizar directly above Kochab, the offset would move equally to the right. These two positions could be bisected to give true north. (If the 12-hour wait moved the observation into daylight, it would mean waiting several months to observe the opposite configuration at night.)

Spence’s interpretation requires that the determinations of north take place at just one of the two vertical alignments of Kochab

and Mizar, generally with Mizar above. Two pyramids that do not conform to this pattern — the second great pyramid at Giza (Khafre) and the later pyramid of Sahure — can easily be explained if the north alignment was done with Kochab above Mizar, producing an error to the west instead of the east.

Spence argues that there was an elaborate 'fixing the north' ceremony that took place early in a new pharaoh's reign, rather than a systematic sequence of observations over months or years. It is an interesting idea, but one that needs to be tested. Can Egyptologists find references to such a ceremony in the ancient texts? This evidence would seem to be crucial for the ultimate acceptance of Spence's ingenious hypothesis, but sadly there is no documentation relating to the construction of the ancient pyramids.

In general, little is known about early

Egyptian astronomy, and even the constellations recorded on the ceilings of tombs remain for the most part unknown and unmatched with the actual starry sky. One of the few identified constellations is the Egyptian adze, the sculptor's mythically powerful tool for making magical images; the adze matches the modern Big Dipper. A pair of such images are depicted in the wall mural of Tutankhamon's tomb — probably the Big and Little Dippers — and elsewhere there is a text about two sharp claws chasing each other around the pole. Could this be an echo of Kochab and Mizar making their alignment rounds? Be on the lookout, Egyptologists, for any such obscure hints.

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.
e-mail: ginger@cfa.harvard.edu

Speciation

Fish found *in flagrante delicto*

Mark Kirkpatrick

Genetic analysis of cichlid fish in Nicaraguan lakes reveals a possible case of repeated sympatric speciation: the creation of two species from one in the same environment.

A species of vertebrate gives rise to another species on average once every few million years. That is somewhat longer than the time span of a typical research grant, which is one reason why speciation is a tough subject to study. But sometimes the process can be caught in the act, giving us a window into the origin of biodiversity.

A paper by Wilson *et al.*¹, just published in *Proceedings of the Royal Society*, reports on just such an apparent case in cichlid fishes

from Nicaragua. The contribution is all the more interesting because it seems that sympatric speciation is taking place. This is the controversial hypothesis which holds that one species can split into two without the benefit of geographical barriers to prevent interbreeding.

The celebrated evolutionist Ernst Mayr built a persuasive case that speciation can occur only when geographical barriers enforce non-random mating while new

species are emerging². This is allopatric speciation, and Mayr's view became the orthodoxy for many years. There is now growing evidence, however, that sympatric speciation can happen in the right circumstances³. The study by Wilson *et al.* breaks new ground by providing the first example of replicated sympatric speciation: one species splitting in two several times independently. If confirmed, these natural replicates would offer a unique opportunity to identify the factors that trigger sympatric speciation.

The cichlid fishes are particularly attractive for speciation studies. The 700 species are remarkable for their diverse colours, shapes and sizes, and behaviours (which include complex courtship and parental care). Cichlids can speciate rapidly: the 300 species in Lake Victoria, east Africa, may have descended from a single ancestral cichlid in only the past 12,400 years⁴. Further, these fish may be prone to sympatric speciation. A remarkable case involves cichlids from isolated crater lakes in Cameroon⁵. Genetic data show that nine species living in Lake Bermin are each other's closest relatives, implying that they speciated there. The lake is only 0.6 km² in surface area and 14.5 m deep, making it implausible that geographical barriers were involved.

Wilson *et al.*¹ studied cichlids in four lakes in Nicaragua. Each lake has two types of fish that differ in colour — a 'normal' morph and a 'gold' morph. In two of the four lakes, Wilson *et al.* found statistically significant differences between the nuclear and/or mitochondrial gene frequencies of the morphs. Further, mitochondria from the different morphs within each lake are statistically more similar to each other than to those from different lakes, supporting the authors' interpretation that sympatric speci-

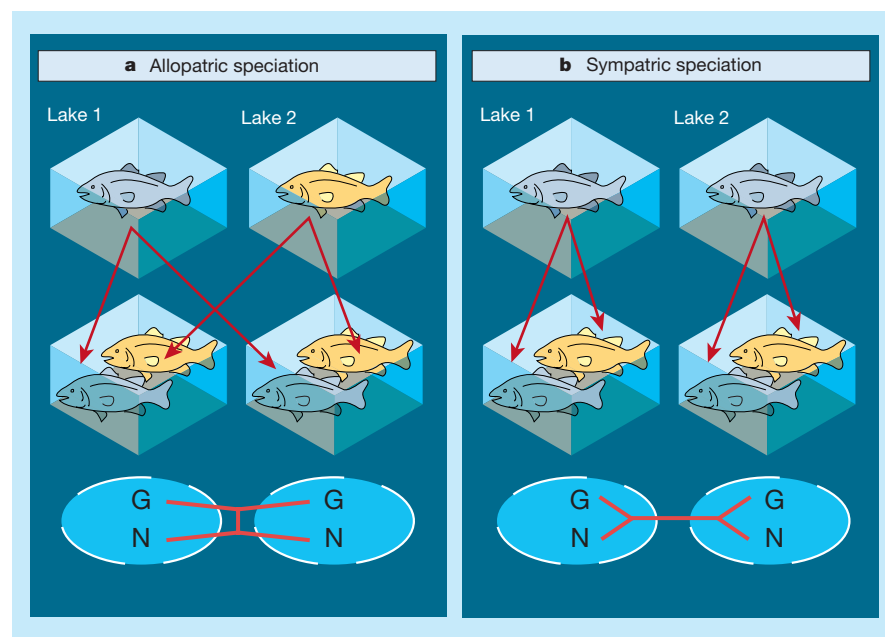


Figure 1 Molecular signatures of speciation: possible outcomes of genetic analyses of cichlid fish in two Nicaraguan lakes.

a, Allopatric speciation, in which species form while a geographical barrier is present and then come into contact secondarily. b, The corresponding pictures for 'replicated sympatric speciation', where new species form independently within each lake in the absence of geographical barriers. At the bottom is the gene tree for genes from gold (G) and normal (N) individuals sampled from the two lakes that would be consistent with each course of events. The gene trees reflect the historical relations of the populations. Under allopatric speciation (a), genes from the gold forms in different lakes will be more similar than genes from the gold and normal forms within each lake. The opposite is true under sympatric speciation (b). Wilson and colleagues' data¹ tend to support this second pattern, and the view that sympatric speciation has occurred.

ation has occurred (Fig. 1). Other explanations are that allopatric speciation occurred on a local scale within each lake; or that parallel invasions of the lakes by two species that originated in different lakes were followed by hybridization between the two, and the spread of mitochondria from one species to the other. These alternatives may be less likely, but they serve to show how hard it is to prove sympatric speciation conclusively, even with solid molecular data.

Earlier studies of these fish⁶ found strong 'assortative' mating based on colour (that is, like-colour morphs mate with like). Population-genetic models show that assortative mating is a particularly powerful mechanism for generating the genetic differences that lead to speciation⁷, so this propensity may have set the stage for sympatric speciation. Even one successful hybridization per generation should prevent substantial divergence for the molecular markers studied by Wilson *et al.*⁸, so apparently there have been strong barriers to genetic mixing between the colour morphs for quite some time.

There is another striking form of variation found in these fish. Cichlids have distinctive bones in their throats called pharyngeal jaws, which have played a key role in their spectacular ecological diversification⁹. The Nicaraguan cichlids have two forms of these jaws, one suited to a diet of snails and the other to soft prey. Typically, speciation involves genetic divergence for ecological traits, and so one would expect one of the two forms of pharyngeal jaws to be strongly associated with each of the colour morphs. They are not. Although there is a correlation, it is only weak (about $r = 0.48$). Further, Wilson *et al.* could not find molecular differences between the pharyngeal morphs.

What gives? Two possibilities come to mind. The first one, favoured by the authors, relates to mathematical models of sympatric speciation^{10,11}. Consider a population that is under divergent selection for two different ecological forms, say eating snails and eating worms, and that also mates assortatively, say with respect to colour. The models show that genetic variation for the ecological trait and for colour will become correlated. This correlation further reinforces divergence of both the ecological trait and colour, and under the right conditions it can cause the population to split in two completely within only a few dozen generations. The Nicaraguan cichlids could be at an intermediate step in this process, in which case the colours, pharyngeal jaws and molecular variation will eventually become tightly associated. A difficulty with that suggestion, however, is that the molecular data suggest the colour morphs have already been isolated for a considerable time.

There is another possibility: that the dramatic variation in pharyngeal jaws may not have genetic causes. The pharyngeal jaws of

these fish change their form in response to diet¹². The two colour morphs live in different habitats at certain times of the year, where different prey are available. Perhaps diet alone directly causes the variation in pharyngeal jaws and its modest correlation with colour. In that case, the jaws may be but an interesting sideshow to speciation.

Non-genetic effects might have a hand in yet another part of this story. In these cichlids, the young spend several weeks in the care of both parents, giving offspring the opportunity to learn the colour of their parents. Perhaps the mating preferences of the fish are influenced in part by behavioural imprinting. Indeed, lab experiments on these cichlids suggest that a fish's response to gold and normal individuals is not affected by its own colour, but is affected by the colour of the parents that reared it¹³. Imprinting may influence speciation in birds¹⁴, so there might be a connection between two remark-

able features of cichlids — their high rates of speciation and their parental care. ■

Mark Kirkpatrick is in the Section of Integrative Biology, University of Texas, Austin, Texas 78712, USA.

e-mail: kirkp@mail.utexas.edu

1. Wilson, A. B., Noack-Kunmann, K. & Meyer, A. *Proc. R. Soc. Lond. B* **267**, 2133–2141 (2000).
2. Mayr, E. *Animal Species and Evolution* (Harvard Univ. Press, 1963).
3. Howard, D. J. & Berlocher, S. H. (eds) *Endless Forms: Species and Speciation* (Oxford Univ. Press, 1998).
4. Johnson, T. C. *et al. Science* **273**, 1091–1093 (1996).
5. Schlieven, U. K., Tautz, D. & Pääbo, S. *Nature* **368**, 629–632 (1994).
6. McKaye, K. R. *Envir. Biol. Fishes* **5**, 75–78 (1980).
7. Kirkpatrick, M. & Ravigné, V. *Am. Nat.* (in the press).
8. Wright, S. *Genetics* **16**, 97–159 (1931).
9. Liem, K. F. *Syst. Zool.* **22**, 425–441 (1973).
10. Dieckmann, U. & Doebeli, M. *Nature* **400**, 354–357 (1999).
11. Kondrashov, A. S. & Kondrashov, F. A. *Nature* **400**, 351–354 (1999).
12. Meyer, A. *Biol. J. Linn. Soc.* **39**, 279–299 (1990).
13. Barlow, G. W., Francis, R. C. & Baumgartner, J. V. *Anim. Behav.* **40**, 713–722 (1990).
14. Irwin, D. E. & Price, T. *Heredity* **82**, 347–354 (1999).

Nanothermodynamics

Breathing life into an old model

Tom Giebultowicz

A classic theory of magnetism has been modernized by a novel use of thermodynamics. The theory can now describe the behaviour of ferromagnetic materials at higher temperatures.

The mean-field theory of ferromagnetism, put forward in 1907 by Pierre Weiss¹, was an important milestone in the development of modern physics. For many years, the Weiss theory was the 'Standard Model' of magnetism. Apart from describing the magnetic properties of materials such as iron, it stimulated great interest in the study of interacting systems, which had an impact far beyond magnetic research.

The theory is still widely used today, even though its predictions are not accurate at certain temperatures. Because of its popularity, attempts to improve the theory's predictions are still worthwhile. One such refinement is reported on page 337 of this issue by Ralph Chamberlin². By including the effects of small-system thermodynamics (nanothermodynamics) to describe local magnetic fluctuations, Chamberlin has extended the mean-field approach to cover the behaviour of ferromagnets at higher temperatures.

The problem tackled by Chamberlin has a long history. Work towards quantitative models of magnetism began in the late 1800s, shortly after classical electromagnetic theory was formulated by J. C. Maxwell. The first successes were in studies of paramagnets and diamagnets — materials that are weakly magnetic under an applied magnetic field (and the magnetization is aligned with the applied field or opposing it, respectively). In

1895, Pierre Curie derived a magnetization law for paramagnets from experiments. Paul Langévin, Curie's pupil, later showed that the law could be derived theoretically by treating a paramagnet as a system of classical non-interacting magnetic dipoles. He also managed to explain diamagnetism using classical electromagnetic theory. But this theory completely failed in its attack on ferromagnets — the materials in which magnetism is strongest. In ferromagnets a sizable magnetization can remain in the material even when there is no external field.

Ferromagnetism abruptly disappears from a material at a certain critical temperature called the Curie point, T_C . Above this, the system behaves as a paramagnet until it is cooled back below T_C and ferromagnetism is restored. Classical physics can explain such effects only if T_C values are at least 1,000 times lower than those observed. Pierre Weiss solved this puzzle by postulating the existence of anomalously strong interactions between the atomic magnetic dipoles. He introduced the interactions into his calculations as a homogenized 'molecular field' proportional to the magnetization. The new model was a spectacular success, qualitatively describing observations at all temperatures.

In particular, for temperatures above T_C , Weiss's model provides an equation for the magnetic susceptibility χ , so that $\chi(T) \propto C/(T - \Theta)$, where C is the Curie

constant and Θ is the Weiss temperature (the model predicts that $\Theta = T_C$). This is the famous Curie–Weiss law and is closely obeyed by all ferromagnets, except at temperatures close to T_C . The theory was primarily the product of Weiss's incredible intuition. In 1907 he could not know the origin of the 'molecular field' he postulated, which only became clear 25 years later, when Heisenberg discovered 'exchange interactions' between atomic spins — forces of a purely quantum nature.

The simplicity of Weiss's concept was an advantage, but also a weakness. In his model each magnetic atom in the sample 'sees' an identical, homogenized environment. It ignores local fluctuations and wave-like excitations that play a crucial role in magnetic behaviour. Consequently, the model's performance is not always satisfactory, especially for $T \rightarrow 0$ (because of wave-like excitations) and around T_C (because of local fluctuations). For instance, the Curie–Weiss law tells us that for temperatures above T_C , where paramagnetic behaviour emerges, a plot of inverse susceptibility, $1/\chi$, against T should be a straight line. This is not the case for temperatures starting 20–50% above T_C , and in experiments T_C is lower than Θ . Also, the model predicts that ferromagnetic ordering completely disappears above T_C . In reality, significant short-range-order effects persist well above T_C through the formation of nanoscopic magnetized regions (clusters) with fluctuating size.

With the development of modern experimental tools, such as NMR spectroscopy and neutron scattering, these deficiencies became clearer. In the 1960s and 1970s, powerful new theories emerged that were better at describing the critical behaviour of ferromagnets near T_C (most notably, renormalization group theory³). In those days many experts believed that the usefulness of the mean-field model had been exhausted and that, like the Bohr model of the hydrogen atom, it should be consigned to the history books. But this did not happen. None of the newer models offer the same versatility as the mean-field theory: each model focuses on effects occurring at different temperatures. Moreover, their underlying formalisms are incompatible, so they cannot be united into one comprehensive picture. The mean-field model is still the only one that describes the qualitative behaviour of ferromagnets over the entire temperature range from $T \rightarrow 0$ to $T \rightarrow \infty$. For that reason, mean-field theory offers a common platform for researchers working on different types of interacting systems — one can say it has become the *lingua franca* of the community.

In his paper, Chamberlin² uses the mean-field formalism to address the entire paramagnetic regime, including temperatures near T_C where significant clustering effects occur. Several ways of incorporating such

effects into the mean-field formalism have been proposed in the past. But previous work considered only fixed-size clusters. Imposing fixed sizes on internal fluctuations was a drastic step, used to make the models tractable at the expense of physical reality. Chamberlin shows that the mathematical difficulties can be overcome by using a thermodynamical formalism known as the 'generalized ensemble'. This was first introduced by Guggenheim⁴ and later used by Hill⁵ in his development of nanothermodynamics. The ensemble allows unrestricted fluctuations in all extensive thermodynamic parameters, including size.

Chamberlin's model gives results for $\chi(T)$ that agree very well with data from several different ferromagnets. It also produces realistic results on the temperature dependence of the

local magnetic order above T_C , thereby creating a robust formalism that appears to solve a very old problem. The success of the model shows that nanothermodynamics — of clear importance for dealing with small systems in many areas of nanoscience — is also crucial for describing internal fluctuations in bulk materials. At last we have a unified picture of the paramagnetic behaviour of ferromagnetic materials. ■

Tom Giebultowicz is in the Department of Physics, Oregon State University, Corvallis, Oregon 97331-6507, USA.

e-mail: tgiebult@PHYSICS.orst.edu

1. Weiss, P. *J. Phys.* **6**, 661–690 (1907).
2. Chamberlin, R. V. *Nature* **408**, 337–339 (2000).
3. Stanley, H. E. *Rev. Mod. Phys.* **71**, S358–S366 (1999).
4. Guggenheim, E. A. *J. Chem. Phys.* **7**, 103 (1939).
5. Hill, T. L. *Thermodynamics of Small Systems (Parts I and II)* (Dover, New York, 1994).

Global change

It's not a gas

Hermann W. Bange

Nitrous oxide is a greenhouse gas and also contributes to ozone loss. It seems that fertilizer run-off into coastal waters stimulates its production — at least to the west of India.

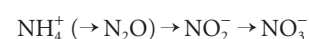
The gas nitrous oxide (N_2O) influences the Earth's climate both directly and indirectly. Occurring at trace levels in the atmosphere, most of the gas comes from only two sources: soils and oceans. But changing environmental conditions, stemming from both natural variability and human-induced perturbations, are increasing nitrous oxide emissions to the atmosphere. On page 346 of this issue, Naqvi *et al.*¹ present the first evidence for a phenomenon that might have far-reaching consequences for the Earth's climate. They show that large amounts of nitrous oxide can temporarily build up in the waters of the Indian continental shelf, and that this is caused by changes in the coastal environment that favour formation of the gas.

In the lower atmosphere (the troposphere) nitrous oxide is present at about only 0.00003%. Nonetheless, it punches well above its weight in terms of its effect on climate and atmospheric chemistry². It is a strong greenhouse gas: a molecule of nitrous oxide is about 200–300 times more powerful in warming potential than a molecule of carbon dioxide. In addition, nitrous oxide is not removed from the troposphere by chemical reactions and so can reach the middle atmosphere (the stratosphere). Here it is photochemically destroyed, forming nitric oxide radicals which are involved in one of the cycles that destroy ozone. So, indirectly, nitrous oxide also affects the stratospheric ozone layer which protects us from harmful ultraviolet radiation.

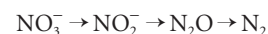
For the past 100 years, the atmospheric concentrations of nitrous oxide have been increasing steadily³. Besides the soils and oceans⁴, there are anthropogenic sources of nitrous oxide such as biomass burning and emissions from industrial processes and automobiles. But the increase of the past century is probably mainly due to the intensification of agriculture⁵: nitrogen-containing fertilizers are known to increase nitrous oxide release from soils⁶. But how the coastal zone responds to increasing inputs of nutrients, in the form of fertilizer run-off from land, remains largely a mystery.

Both the terrestrial and the oceanic nitrogen cycles are driven by the activities of microorganisms. Some of them can obtain their energy by transforming various forms of nutrients such as nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+). Nitrous oxide is mainly produced during two processes: nitrification and denitrification.

In nitrification, ammonium is transformed into nitrate. Nitrous oxide is formed as a by-product during the step from ammonium to nitrite:



Denitrification occurs when nitrate is transformed into dinitrogen (N_2):



But the formation of dinitrogen can be incomplete, generating nitrous oxide —

estimates vary, but it seems that a yield of 0.1–6% of nitrous oxide can be produced.

But things are more complex than this because the amount of nitrous oxide produced by either nitrification or denitrification depends on the prevailing oxygen conditions — maximum yields of the gas occur only in a narrow range of low oxygen concentrations. This means that significant nitrous oxide accumulations in the ocean will probably mainly be found under oxygen-depleted (suboxic) conditions⁷. In turn, this implies that small variations in the levels of dissolved oxygen might lead to large changes in the rate at which nitrous oxide is formed.

Naqvi *et al.*¹ provide evidence for a dramatic decrease in dissolved oxygen in the coastal waters off India during the monsoon season, which in turn is associated with the large-scale formation of nitrous oxide. It seems that fertilizer run-off intensifies oxygen depletion, by increasing the productivity of marine algae and the subsequent recycling of the freshly produced organic matter. We should be concerned about these observations because seasonal oxygen depletion seems to have developed only over the past few years, and it is not locally restricted — it spreads along the whole Indian continental shelf.

A rough calculation underlines this point. Naqvi *et al.* estimate that the nitrous oxide emissions from the area of ocean that they investigated over a period of only six months range from 0.06×10^{12} g to 0.39×10^{12} g. Comparing those figures with estimates of the global annual output of nitrous oxide from the oceans^{8,9} (between 1.9×10^{12} g and 17×10^{12} g) reveals a contribution of 0.4–21% from an area which is only about 0.05% of that of the world's

oceans. From Naqvi and colleagues' data, it seems likely that increasing nutrient inputs into the coastal zone from nitrogen fertilizers cause the seasonal shift in oxygen regime. But natural variability stemming from mechanisms such as the monsoon-induced upwelling of coastal waters, which delivers nutrients to the surface, cannot yet be excluded.

Several questions remain. Are there comparable accumulations of nitrous oxide in other oxygen-depleted coastal environments (for instance the Gulf of Mexico)? Are such accumulations a short- or a long-term phenomenon, and do they affect the atmospheric concentrations of the gas? Answers might come from intensive time-series measurements in selected coastal areas. Meantime, plenty of readers will have recognized nitrous oxide in a more familiar guise as laughing gas. But the results of Naqvi *et al.*

show that, in environmental terms, this may be no laughing matter.

Hermann W. Bange is in the Biogeochemistry Department, Max Planck Institute for Chemistry, PO Box 3060, 55020 Mainz, Germany.
e-mail: bange@mpch-mainz.mpg.de

1. Naqvi, S. W. A. *et al.* *Nature* **408**, 346–349 (2000).
2. Prather, M. *et al.* in *Climate Change 1995: The Science of Climate Change, Contribution of Working Group I to the Second Assessment of the Intergovernmental Panel on Climate Change* (eds Houghton, J. T. *et al.*) 86–103 (Cambridge Univ. Press, 1996).
3. Battle, M. *et al.* *Nature* **383**, 231–235 (1996).
4. Bouwman, A. F., Van der Hoek, K. W. & Olivier, J. G. J. *J. Geophys. Res.* **100**, 2785–2800 (1995).
5. Kroeze, C., Mosier, A. & Bouwman, L. *Glob. Biogeochem. Cycles* **13**, 1–8 (1999).
6. Robertson, G. P., Paul, E. A. & Harwood, R. R. *Science* **289**, 1922–1925 (2000).
7. Codispoti, L. A. *et al.* in *Oceanography of the Indian Ocean* (ed. Desai, B. N.) 271–284 (Balkema, Rotterdam, 1992).
8. Nevison, C. D., Weiss, R. F. & Erickson, D. J. III *J. Geophys. Res.* **100**, 15809–15820 (1995).
9. Bange, H. W., Rapsomanikis, S. & Andreae, M. O. *Glob. Biogeochem. Cycles* **10**, 197–207 (1996).

Evolutionary biology

Mitochondrial genes on the move

Michael W. Gray

In flowering plants, genes have frequently been transferred from mitochondria to the cell nucleus by way of a remarkable evolutionary rapid-transit system.

Mitochondria, the cell's energy-generating organelles, have their own genome, known as mitochondrial (mt) DNA. This genome is, however, minuscule compared with that of the free-living bacterium from which mitochondria originated¹ in eukaryotes — organisms whose

cells have a defined nucleus. Gene transfer from the mitochondrial to the nuclear genome has played a central role in this process of reductive genome evolution². The protein product of such a transferred gene still functions in the mitochondrion, but the gene instead operates from the nucleus.

Medicine

Cardiac arrest can be less of a gamble

There is no good place to suffer cardiac arrest. But if you do, a hospital is obviously one location where quick treatment is likely to increase the chances of recovery. As described in two papers in *The New England Journal of Medicine*, certain casinos in Nevada and Mississippi, and some airlines, also make recovery a better bet.

The type of cardiac arrest concerned is that caused by ventricular fibrillation, in which contraction of heart-muscle fibres becomes uncoordinated. The key to treatment is the use of defibrillating equipment, and swiftly: with each minute that passes, the possibility of recovery falls rapidly.

T. D. Valenzuela *et al.* (*N. Engl. J. Med.* **343**, 1206–1209; 2000) report

the results of a programme in which defibrillators were installed in casinos and security staff were trained to use them. The overall success rate, judged by a patient's eventual discharge from hospital, was 53%. The national average is just 5%.

The other study (R. L. Page *et al.* **343**, 1210–1216; 2000) involved American Airlines. Like some other aircraft operators, American Airlines has introduced defibrillators on its planes and trained flight attendants to use them. Here the rate of survival was 40%, which compares well even with the rates achieved by quick-response emergency services in other circumstances. The authors calculate that if all commercial aircraft carried the equipment, 93 people would survive who would



otherwise have died each year.

The authors also point to a study showing that age need not be a barrier to the effective use of defibrillators. Twelve-year-old children took only 27 seconds longer to carry out defibrillation than medical personnel, albeit in simulations.

But perhaps a safer approach is to avoid flying or gambling altogether. Cardiac arrest is more common in these situations than in most other public places — although, in the case of casinos, it is not clear whether the main factor is winning, losing or just being there.

Tim Lincoln

Astronomy

The legacy of a lonely life

Neutron stars are extremely dense, compact stars consisting mostly of neutrons. They are thought to be the hot remains of a supernova explosion — when an old massive star collapses in on itself. Most known neutron stars are either pulsating (pulsars) or have a close companion.

There are supposed to be as many as a billion lone neutron stars zooming about the Milky Way, but the first isolated, non-pulsating neutron star was discovered just five years ago by Fred Walter of the State University of New York at Stony Brook (*Nature* **379**, 233–235; 1996). The initial X-ray discovery was confirmed by optical data from the Hubble Space Telescope (*Nature* **389**, 358–360; 1997). Last week, at a meeting of the American Astronomical Society in

Honolulu, Hawaii, Walter revealed the latest Hubble observations of this extraordinary object (*The Astrophysical Journal*, in the press).

The extreme densities of neutron stars offer a natural laboratory for exploring the 'equation of state' of nuclear matter — the relationship between pressure, temperature and radius in the atomic nucleus. Although the first optical measurements of the neutron star gave Walter a fairly good idea of its radius and distance, he needed to know its parallax to be sure. Parallax is the apparent motion of nearby objects, when seen against the background of more distant ones. As the Earth orbits the Sun, the positions of nearby stars shift slightly, with a maximum every six months.



From observations over three years (see image), Walter calculated a distance of 200 light years for the star, making it the closest known neutron star to Earth. He found its radius to be just 11 km; together with the mass this places constraints on the nuclear equation of state. Walter's results also show that the star is speeding away (at about 108 km s⁻¹) from a group of stars known

as the Upper Sco OB association, from which it was probably ejected a million or so years ago. When combined with the X-ray and optical observations, this date allows us to estimate how quickly neutron stars cool, which for this star seems to be faster than theoretical expectations. Quite a legacy for what was once just another mystery X-ray source.

Leslie Sage

NASA/F. M. WALTER

On page 354 of this issue, Adams *et al.*³ pose the question, "During the evolution of a particular eukaryotic lineage, how many times and how often has a given mitochondrial gene been relocated from organelle to nucleus?". In the case of angiosperms (flowering plants), the answer is, "Very many times and very often, indeed".

Even the most gene-rich mtDNAs encode at least ten times fewer genes than the smallest bacterial genomes. So most of the reduction in mitochondrial genome size and gene content is assumed to have occurred early in mtDNA evolution. However, the extremely variable gene content of contemporary mitochondrial genomes implies that mitochondrial gene loss has been an on-going process. By mapping gene losses onto phylogenetic trees, we can infer that the same mitochondrial gene has been transferred to the nucleus in separate events in different eukaryotic lineages⁴. Individual instances of mitochondrion-to-nucleus gene transfer have been well documented⁵. Up to now, however, we have had little appreciation of the overall frequency and timing of such gene relocations during mtDNA evolution.

Adams *et al.*³ report the remarkably frequent loss from angiosperm mtDNA of the gene encoding ribosomal protein S10 (a component of the mitochondrial ribosome) and the transfer of this gene (*rps10*) to the nucleus. Their results conjure up a picture of an evolutionary rapid-transit system, in which *rps10* is being continually and frequently transported from the mitochondrial to the nuclear genome in different groups of

angiosperms — and at a dizzying pace in evolutionary terms.

The authors used part of the *rps10* gene as a hybridization probe to score the presence or absence of this gene in DNA extracted from 277 angiosperms representing 169 families. The lab concerned has creatively exploited this valuable resource in other studies^{6,7}. The approach works well in plants for two reasons. First, there are many copies of the mitochondrial genome. Second, angiosperm mitochondrial genes evolve very slowly compared with single-copy nuclear genes — not to mention their counterparts in the mitochondria of animals, fungi and protists.

When plotted onto a phylogenetic tree of the angiosperms examined, the hybridization results indicated a loss of *rps10* on 26 separate occasions. In 16 of these cases, the authors amplified a nuclear *rps10* counterpart (not detectable by hybridization) using the polymerase chain reaction. Because a single mitochondrion-to-nucleus transfer of *rps10* could conceivably be followed by many independent losses of this gene from the mitochondrial genome, several of the nuclear *rps10* genes were characterized in more detail to see how many separate *rps10* transfers these 16 losses may represent.

To become active in the nucleus, a gene acquired from mtDNA must be inserted into the nuclear genome in such a way that a mature, translatable messenger RNA can be produced. Moreover, the resulting protein (made outside the nucleus, in the cell's cytosol) must be targeted to and imported into mitochondria. What emerges from the

analysis of such transferred mitochondrial *rps10* genes is an intriguing diversity in the mechanism of functional activation in different angiosperms.

In some cases, pre-existing copies of other nuclear genes have been parasitized, with the *rps10* coding sequence being inserted into the host gene. In several instances, a mitochondrial targeting sequence (pre-sequence) has been effectively co-opted to provide entry for the RPS10 protein back into the mitochondrion; however, different nuclear genes are used as the source of the presequence in different plants. In other cases, the nucleus-encoded RPS10 protein appears to be imported into mitochondria despite the absence of an obvious pre-sequence. From these results and other published data, the authors infer that there have been at least seven separate transfers of *rps10* to the nucleus in angiosperms, and probably many more.

Is this just the tip of the iceberg? Apparently it is. Adams *et al.*³ point out that their study encompasses only about 2% of living angiosperm genera, so that *rps10* may have been "independently transferred hundreds of times during angiosperm evolution". Other genes that encode ribosomal proteins also seem to have been lost frequently from angiosperm mitochondrial genomes. Individual examples of such transfers have been characterized, again illuminating the myriad mechanisms that seem to be employed for functional activation. In grasses, for instance, a relocated *rps14* gene has ended up within a nuclear gene (*sdhB*) that encodes a subunit of the mitochondrial enzyme succi-

nate dehydrogenase. Co-expression of the *sdhB* and *rps14* coding sequences, followed by alternative splicing of the resulting mRNA, gives rise to two different proteins having the same mitochondrial targeting signal^{8,9}.

Further studies have yielded evidence of comparably high and variable rates of loss of a majority of 13 other mtDNA-encoded ribosomal protein genes during angiosperm evolution (K. Adams, Y.-L. Qiu and J. D. Palmer, personal communication). In contrast, only two of a sample of 13 genes involved in energy generation showed a similarly high rate of loss from angiosperm mtDNA. A particularly striking point is that in many plants the relative rate of ribosomal-protein gene loss actually exceeds the rate of nucleotide substitution at a single silent site (a site at which a mutation does not result in an amino-acid change): this is a conclusion that Adams *et al.* (and I) find astonishing.

In other eukaryotes (animals, for instance), further transfer of genes from mitochondria to the nucleus has effectively been blocked by changes in the mitochondrial genetic code: essentially, mitochondrial gene content has been frozen. In plants, mitochondrion-to-nucleus gene relocation is undoubtedly facilitated by the fact that there is no genetic code

barrier to such transfer: the systems for translating RNA into protein use the same genetic code in both mitochondria and the cytosol. Even so, we could hardly have imagined the relative ease with which angiosperms seem to be able to shuttle mitochondrial genes into the nucleus and activate them there. Unravelling the genetic and biochemical workings of this highly efficient rapid-transit system will undoubtedly provide further surprises down the road. ■

Michael W. Gray is in the Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.

e-mail: M.W.Gray@Dal.Ca

1. Gray, M. W., Burger, G. & Lang, B. F. *Science* **283**, 1476–1481 (1999).
2. Berg, O. G. & Kurland, C. G. *Mol. Biol. Evol.* **17**, 951–961 (2000).
3. Adams, K. L., Daley, D. O., Qiu, Y.-L., Whelan, J. & Palmer, J. D. *Nature* **408**, 354–357 (2000).
4. Gray, M. W. *Curr. Opin. Genet. Dev.* **9**, 678–687 (1999).
5. Schuster, W. & Brennicke, A. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 61–78 (1994).
6. Cho, Y., Qiu, Y.-L., Kuhlman, P. & Palmer, J. D. *Proc. Natl Acad. Sci. USA* **95**, 14244–14249 (1998).
7. Qiu, Y.-L., Cho, Y., Cox, J. C. & Palmer, J. D. *Nature* **394**, 671–674 (1998).
8. Figuerosa, P., Gómez, I., Holuigue, L., Araya, A. & Jordana, X. *Plant J.* **18**, 601–609 (1999).
9. Kubo, N., Harada, K., Hirai, A. & Kadowaki, K.-i. *Proc. Natl Acad. Sci. USA* **96**, 9207–9211 (1999).

Neural engineering

Real brains for real robots

Sandro Mussa-Ivaldi

Neural signals from the brains of monkeys have been used to drive the movement of robotic arms. The ultimate objective of such work is to design controllable prosthetic limbs.

The idea of driving robotic limbs with what effectively amounts to the mere 'power of thought' was once in the realm of science fiction. But this goal is edging closer to reality, thanks in part to two decades of studies that have revealed a close match between the activity of neurons in the brain's cerebral cortex and the movements of the hand¹. Now, writing on page 361 of this issue², Wessberg and colleagues describe how they have used electrical signals from five regions in the cerebral cortex of monkeys to drive the movement of robotic arms.

Such 'brain-computer interfacing'³ has clear-cut clinical aims. The ambitious goal is to provide amputees and patients suffering from a variety of severe motor disorders — such as paralysis and amyotrophic lateral sclerosis — with the means to act and communicate by replacing the control of muscles with the control of artificial devices by brain activity.

The idea of using signals from the cerebral cortex to drive artificial limbs was pioneered 30 years ago⁴, when these signals were

used to predict the movement of the wrist in real time. But only since the motor cortex region of the brain has been studied in detail⁵ have researchers begun to explore the possibility of reproducing the full complexity of arm movements⁶. To do so one requires the ability to extract — in real time — the information concerning arm movements that is hidden within the activities of large groups of neurons⁷.

Researchers are gradually developing the hardware and software needed to connect brains to robotic limbs. The hardware includes 'conic electrodes'⁸, an innovative development based on the idea of inducing nerve growth within small, hollow electrodes containing nerve growth factor. This technique establishes stable, long-lasting electrical contacts between brain cells and electrodes.

Wessberg *et al.*², meanwhile, have been working on the software. They provide a simple solution to the problem of extracting 'movement information' from neural signals detected by microelectrodes implanted in



100 YEARS AGO

A simple method of recording the speed of motor cars and other vehicles has been devised by M. L. Gaumont, and accounts of the device appear in *Cosmos* and *La Nature* of November 3. The instrument consists simply of a camera with a double shutter, by which two exposures are made of the same plate, separated by a known interval of time. On developing the photograph, two images are obtained of the moving object, and, by measuring the distance between them, the dimensions of the car being supposed known and also measured on the plate, it is easy to calculate the speed of the car at the instant when the photograph was taken. The object is to assist the authorities in regulating the speed of these vehicles and checking furious driving.

From *Nature* 15 November 1900.

50 YEARS AGO

In 1925, Prof. Dart announced the discovery of a new type of higher primate that seemed to be somewhat intermediate between ape and man. This was the skull of the Taungs child which he called *Australopithecus africanus*. For some years there was considerable dispute between those of us who regarded the little skull as that of a being closely allied to the ancestor of man, and those who considered it only a variety of ape, allied to the chimpanzee, which by parallelism had come to resemble man in many characters. Since 1936 we have made a large number of new discoveries, and this group of higher primates is now known by many skulls and skeletal remains of adults of a considerable variety of forms which we think should be placed in different genera and species. In the past two years we have discovered a number of nearly complete skulls which give us a new picture of the origin of man. When the only known skulls of our so-called 'ape-men' had brains of between 450 and 650 c.c., those who said they were only apes and had no close relationship to man seemed to have some case. But now that we have skulls with brains of 750, 800 and possibly 1,000 c.c., we seem to be dealing with beings that have some claim to be human...Possibly we are correct in assuming that there lived in South Africa a million or perhaps two million years ago a family of higher primates, not closely related to the living anthropoids, but perhaps evolved from a very early anthropoid or even a pre-anthropoid by a different line...

From *Nature* 18 November 1950.

different regions of the cerebral cortex of living owl monkeys (Fig. 1). The neural signals determine the next position to be assumed by an artificial arm. At each instant, a computer program developed by Wessberg *et al.*, running in real time, determines the next position of the artificial hand on the basis of the neural signals collected during the past second. As this process is continuously repeated, the program generates a sequence of hand positions to create an entire trajectory of hand movement.

In this computer program, the relationship between neural firing and hand position is of the simplest linear kind. The activity of each neuron — measured in impulses per second — is multiplied by a numerical coefficient, and the outcome is added to the outcomes from other neurons. The authors determined the coefficients at the beginning of the experiment by matching neural activities with the movements of each monkey's own hand. Wessberg *et al.* tried several computational schemes of varying complexity, using a repertoire of models. Remarkably, the simplest linear approach turned out to perform as well as the more complex methods.

In this way, the authors solve the problem of generating natural movements of a robotic arm in just one direction (left or right) or in three dimensions. They also find that the

coefficients derived from the analysis of natural hand movements towards targets on the right side of the monkey can be used to guide the robot arm in different movements towards targets on the left side. This 'generalization' is an important property of Wessberg *et al.*'s method. It will be important to explore further the ability of this approach to generate movements of the robot arm over a wide region of space, but this work represents a first, innovative step.

However, the creation of interfaces between neural tissue and machines has implications beyond the development of controllable prosthetics. It will also help us to understand how the brain works.

For example, the use of 'neurorobotic' systems may prove invaluable in working out which regions in the cerebral cortex control which features of limb movements. Wessberg *et al.* placed signal-recording microelectrodes in five different areas in the cerebral cortex, including the left primary motor cortex, the posterior parietal cortex, and the premotor cortex of both left and right hemispheres. The dorsal premotor cortex is believed to plan the general spatial and temporal features of movements. The left motor cortex presides over the generation of movement commands for the right arm. And the posterior parietal cortex is thought to integrate visual, sensory and motor infor-

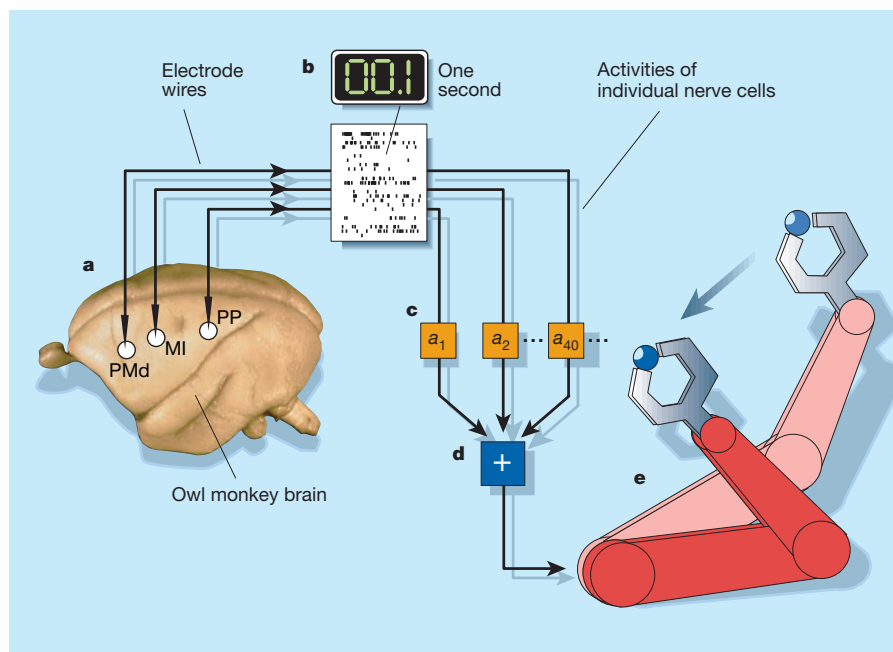


Figure 1 Controlling robotic arms with real neurons. Using live owl monkeys, Wessberg *et al.*² placed several microwire electrodes in different areas of the cerebral cortex that are involved in controlling natural arm movements (a). PMd, dorsal premotor cortex; M1, primary motor cortex; PP, posterior parietal cortex. b, From the signals detected by each electrode, a computer program extracted the activities of individual nerve cells. c, The activity of each cell at each instant of time was multiplied by numerical coefficients (a_1 , a_2 , and so on). d, The results of this operation for all the neurons were added together to determine the position of the robot hand (e). To derive the hand position at a given instant of time, the program combined the neural data collected during the preceding period of about one second. The movements of the robot arm reproduced the natural arm movements made by the monkey during the experiment.

Daedalus

David Jones

David Jones, author of the Daedalus column, is indisposed.

mation to determine the location of a movement target (such as a piece of food), and how to reach it. When the authors analysed the efficiency of their program in extracting arm-movement information from these different areas, they found — somewhat unexpectedly — that the best place to put the electrodes was in the left dorsal premotor cortex. It seems that we still have much to learn about the functions of the cerebral cortex.

In addition, earlier work^{9,10} showed that it is possible to train cortical neurons in monkeys and in rats to control simple devices. Neural populations that would normally activate the muscles of a limb learn to activate an external motor. Remarkably, after learning was completed, these neurons stopped producing the natural movement of the limb — disproving the view that there is a fixed, unchangeable pattern of connections between cortical neurons and muscles.

As we study the brain using the metaphor of computers — and computers using the metaphor of the brain — we face the question of whether and how biological neurons can be 'programmed' to carry out specific computations such as those needed to control a robotic arm. Achieving this goal, which would give us direct access to the unsurpassed computational power of the brain, would represent a huge step forwards. The use of artificial devices to test the learning ability of networks of real neurons may also help us to understand the mechanisms inherent to any form of biological learning, such as long-lasting increases and decreases in signal transmission across neuronal synapses. Harnessing these mechanisms is perhaps the greatest challenge in building prosthetic devices controlled by brain activity.

Sandro Mussa-Ivaldi is in the Departments of Physiology and of Physical Medicine and Rehabilitation, Northwestern University Medical School, and in the Sensory Motor Performance Program, Rehabilitation Institute of Chicago, 303 East Chicago Avenue, Chicago, Illinois 60611, USA. e-mail: sandro@northwestern.edu

1. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. *J. Neurosci.* **2**, 1527–1537 (1982).
2. Wessberg, J. *et al.* *Nature* **408**, 361–365 (2000).
3. Wolpaw, J. *et al.* *IEEE Trans. Rehabil. Eng.* **8**, 164–173 (2000).
4. Humphrey, D., Schmidt, E. & Thompson, W. *Science* **170**, 758–762 (1970).
5. Georgopoulos, A. *Annu. Rev. Neurosci.* **14**, 361–377 (1991).
6. Isaacs, R., Weber, D. & Schwartz, A. *IEEE Trans. Rehabil. Eng.* **8**, 196–198 (2000).
7. Lin, S., Si, J. & Schwartz, A. *Neural Comput.* **9**, 607–621 (1997).
8. Kennedy, P. *J. Neurosci. Meth.* **29**, 181–193 (1989).
9. Fetz, E. & Baker, M. A. *J. Neurophysiol.* **36**, 179–204 (1973).
10. Chapin, J., Moxon, K., Markowitz, R. & Nicolelis, M. *Nature Neurosci.* **2**, 664–670 (1999).

Surfing the p53 network

Bert Vogelstein, David Lane and Arnold J. Levine

The p53 tumour-suppressor gene integrates numerous signals that control cell life and death. As when a highly connected node in the Internet breaks down, the disruption of p53 has severe consequences.

Tumour-suppressor genes are needed to keep cells under control. Just as a car's brakes regulate its speed, properly functioning tumour-suppressor genes act as brakes to the cycle of cell growth, DNA replication and division into two new cells. When these genes fail to function properly, uncontrolled growth — a defining feature of cancer cells — ensues.

The p53 gene, first described in 1979, was the first tumour-suppressor gene to be identified. It was originally believed to be an oncogene — a cell-cycle accelerator (Box 1) — but genetic and functional data obtained ten years after its discovery showed it to be a tumour suppressor. Moreover, it was found that the p53 protein does not function correctly in most human cancers (Fig. 1). In about half of these tumours, p53 is inactivated directly as a result of mutations in the p53 gene. In many others, it is inactivated indirectly through binding to viral proteins, or as a result of alterations in genes whose products interact with p53 or transmit information to or from p53.

The realization that p53 is a common denominator in human cancer has stimulated an avalanche of research since 1989. During that time there have been over 17,000 publications centred on p53 — 3,300 in the past year alone — and over 10,000 tumour-associated mutations in p53 have been discovered, in organisms ranging from humans to clams^{1,2}. As might be expected, this work has led not only to considerable insights into tumour development, but also to consider-

Mechanism of inactivating p53	Typical tumours	Effect of inactivation
Amino-acid-changing mutation in the DNA-binding domain	Colon, breast, lung, bladder, brain, pancreas, stomach, oesophagus and many others	Prevents p53 from binding to specific DNA sequences and activating the adjacent genes
Deletion of the carboxy-terminal domain	Occasional tumours at many different sites	Prevents the formation of tetramers of p53
Multiplication of the MDM2 gene in the genome	Sarcomas, brain	Extra MDM2 stimulates the degradation of p53
Viral infection	Cervix, liver, lymphomas	Products of viral oncogenes bind to and inactivate p53 in the cell, in some cases stimulating p53 degradation
Deletion of the p14 ^{ARF} gene	Breast, brain, lung and others, especially when p53 itself is not mutated	Failure to inhibit MDM2 and keep p53 degradation under control
Mislocalization of p53 to the cytoplasm, outside the nucleus	Breast, neuroblastomas	Lack of p53 function (p53 functions only in the nucleus)

Figure 1 The many ways in which p53 may malfunction in human cancers.

able confusion and controversy. Here we suggest that signalling pathways involving p53 — like cellular signalling pathways in general — cannot be understood by looking at isolated components. Instead, it is essential to consider the tangled networks into which these signalling components are integrated.

Activating the p53 network

The p53 network is normally 'off'. It is activated only when cells are stressed or damaged. Such cells pose a threat to the organism: they are more likely than undamaged cells to contain mutations and exhibit abnormal cell-cycle control, and present a greater risk of becoming cancerous. The p53 protein shuts down the multiplication of stressed cells, inhibiting progress through the cell cycle. In

many cases it even causes the programmed death (apoptosis) of the cells in a desperate attempt to contain the damage and protect the organism. The p53 protein therefore provides a critical brake on tumour development, explaining why it is so often mutated (and thereby inactivated) in cancers.

What sort of stresses, then, activate the p53 network? Early work focused on DNA damage as the 'on' switch. A single break in a double-stranded DNA molecule may be sufficient to trigger a rise in levels of p53 protein. This remarkable sensitivity to DNA damage confounded subsequent studies that sought to establish whether the p53 response could be triggered by other signals. It was difficult to show that these other signals did not cause at least a few breaks in double-stranded

Box 1 The genes that cause cancer

Oncogenes. These are analogous to the accelerators in a car. Oncogenes stimulate appropriate cell growth under normal conditions, as required for the continued turnover and replenishment of the skin, gastrointestinal tract and blood, for example. A mutation in an oncogene is tantamount to having a stuck accelerator: even when the driver releases his foot from the accelerator pedal, the car continues to move. Likewise, cells with mutant oncogenes continue to grow (or refuse to die) even when they are

receiving no growth signals. Examples are Ras, activated in pancreatic and colon cancers, and Bcl-2, activated in lymphoid tumours.

Tumour-suppressor genes. When the accelerator is stuck to the floor, the driver can still stop the car by using the brakes. Cells have brakes, too, called tumour-suppressor genes. These keep cell numbers down, either by inhibiting progress through the cell cycle and thereby preventing cell birth, or by promoting programmed cell death (also called

apoptosis). Just as a car has many brakes (the foot pedal, handbrake and ignition key), so too does each cell. When several of these brakes are rendered non-functional through mutation, the cell becomes malignant. Examples are the gene encoding the retinoblastoma protein, inactivated in retinoblastomas, p53 (Fig. 1), and p16^{INK4a}, which inhibits cyclin-dependent kinases and is inactivated in many different tumours.

Repair genes. Unlike oncogenes and tumour-suppressor genes, repair

genes do not control cell birth or death directly. They simply control the rate of mutation of all genes. When repair genes are mutated, cells acquire mutations in oncogenes and tumour-suppressor genes at an accelerated rate, driving the initiation and progression of tumours. In the car analogy, a defective repair gene is much like having a bad mechanic. Examples are nucleotide-excision-repair genes, whose inactivation leads to susceptibility to skin and colon tumours, respectively.

DNA. Recent research, however, has confirmed the existence of at least three independent pathways by which the p53 network can be activated.

One pathway is indeed triggered by DNA damage, such as that caused by ionizing radiation. Here the activation of the network is dependent on two protein kinases — enzymes that add phosphate groups to other proteins. Two of the major kinases in question are called ATM (for ataxia telangiectasia mutated, named after a disease in which this enzyme is mutated) and Chk2 (ref. 3). ATM is stimulated by double-strand breaks, and Chk2 is in turn stimulated by ATM.

The second pathway is triggered by aberrant growth signals, such as those resulting from the expression of the oncogenes Ras or Myc. In this case, activation of the p53 network in humans depends on a protein called p14^{ARF} (refs 4,5).

The third pathway is induced by a wide range of chemotherapeutic drugs, ultraviolet light, and protein-kinase inhibitors. This pathway is distinguished from the others because it is not dependent on intact ATM, Chk2 or p14^{ARF} genes, and may instead involve kinases called ATR (ataxia telangiectasia related) and casein kinase II⁶.

All three pathways inhibit the degradation of p53 protein, thus stabilizing p53 at a high concentration. The increased concentration of p53 — covalently modified as described below — allows the protein to carry out its major function: to bind to particular DNA sequences and activate the expression (transcription) of adjacent genes. These genes, directly or indirectly, lead ultimately to cell death or the inhibition of cell division — but more on this later.

Stabilizing and modifying p53

The amount of p53 protein in cells is determined mainly by the rate at which it is degraded, rather than the rate at which it is made. The degradation proceeds through a process called ubiquitin-mediated proteolysis. Through a series of steps, several copies of a small peptide (ubiquitin) are attached to the protein to be degraded (in this case p53). This ubiquitin chain acts as a 'flag', enabling p53 to be detected by the protein-degrading machinery. The MDM2 protein is one of the enzymes involved in labelling p53 with ubiquitin⁷.

This process is subject to a feedback loop like those found in electrical circuits. The p53 protein binds to the regulatory region of the MDM2 gene and stimulates the transcription of this gene into messenger RNA, which is then translated into protein. This MDM2 protein then binds to p53 and stimulates the addition of ubiquitin groups to the carboxy terminus (the end) of p53, which is then degraded. This lowers the concentration of p53 and reduces transcription of the MDM2 gene, closing the feedback loop and allowing p53 levels to rise again.

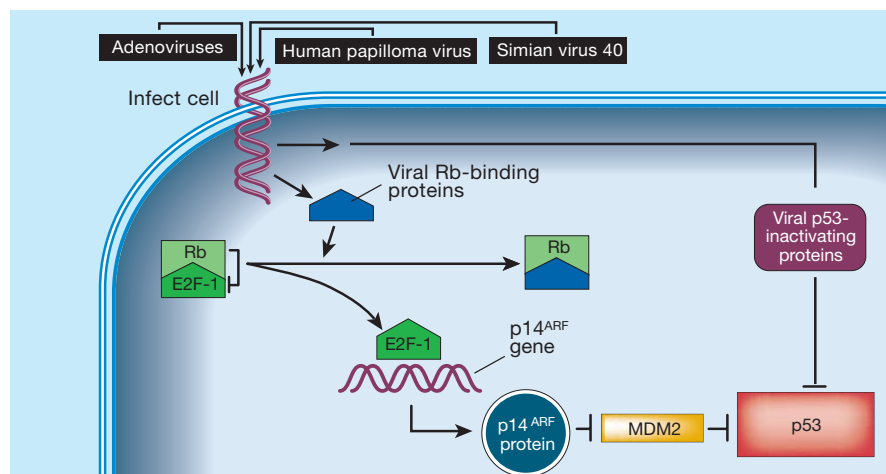


Figure 2 Viral oncogenes and the p53 network. Several viruses encode proteins that block the interaction between an infected cell's retinoblastoma protein (Rb) and transcription factors of the E2F family, such as E2F-1. This frees E2F-1 to activate target genes required for cellular proliferation (not shown). But it also results in the production of the p14^{ARF} protein, interference with the activity of MDM2 (a negative regulator of the p53 protein), and consequent stabilization of p53. This slows cell (and hence viral) replication. The viruses counteract these cellular defences by producing proteins that inhibit the function of p53. This predisposes the infected cells to become cancerous.

But an increased level of cellular p53 protein alone is not sufficient for it to become a transcriptional activator. This requires conformational changes in the protein, resulting from modifications such as the addition or removal of phosphate, acetyl, glycosyl, ribose, ubiquitin or 'sumo' chemical groups^{6,8,9} ('sumo' is a ubiquitin-like polypeptide that can reversibly modify proteins). The carboxy terminus of p53 normally folds back and inhibits the DNA-binding domain located in the central part of the p53 protein. Acetylation of lysine residues or phosphorylation of serine residues near the carboxy terminus of p53 can enhance the binding of p53 to DNA, presumably by interfering with this folding. Interestingly, such conformational changes can also be achieved by antibodies, peptides and drugs that interact with the carboxy terminus¹⁰. These compounds might represent a new way to enhance the function of normal p53 and to restore normal function to mutant p53.

Phosphorylation of the amino terminus (the start) of p53 does not affect its DNA-binding abilities, but does affect its affinity for MDM2 and subsequent degradation. Other changes to the p53 protein and its MDM2 partner are also important in the p53 network. For example, sumolation of MDM2 might reduce its ubiquitination (and hence degradation)¹¹. This would mean that there is more MDM2 around to ubiquitinate p53, so stimulating p53 degradation.

When a protein promotes the synthesis of its own negative regulator, the levels of the two proteins in a cell would be expected to oscillate out of phase with each other. This has been observed for p53 and MDM2 (ref. 12). Similarly, any perturbation of either p53 or MDM2 should have dramatic effects on the other, as well as on the behaviour of cells

and organisms. This is shown by the fact that mice genetically engineered to lack both MDM2 and p53 survive to adulthood, whereas mice lacking only MDM2 die as embryos¹³ — presumably because of the unchallenged activity of p53.

Linking activation to stabilization

The most intensively investigated pathway to p53 activation is the one that is initiated by DNA damage³. This damage is sensed by 'checkpoints' that retard progress through the cell cycle until the damage is mended. The checkpoint proteins that sense and signal DNA damage have been remarkably conserved during evolution, being found in organisms spanning yeast to humans. They include several kinases, particularly DNA-dependent protein kinase, ATM, Chk1 and Chk2 (ref. 3). All four of the mammalian forms of these kinases phosphorylate p53 at amino-terminal sites that are close to the MDM2-binding region of the protein^{6,8,9}. These results have led to a seductive model in which these kinases, activated by DNA damage, phosphorylate the p53 protein and thereby block its interactions with MDM2, leading to stabilization of p53.

But it has been shown that p53 molecules lacking most phosphorylation sites can still be stabilized in response to DNA damage and still activate p53-dependent gene transcription. This suggests that the activation of p53 is not fully controlled by any single phosphorylation site or protein^{6,8,9}. On the other hand, patients with inherited mutations of the Chk2 gene are predisposed to cancer. This syndrome is remarkably similar to that seen in patients with inherited mutations of p53 (ref. 14) — compelling evidence for the importance of checkpoints that sense DNA damage in the p53 network.

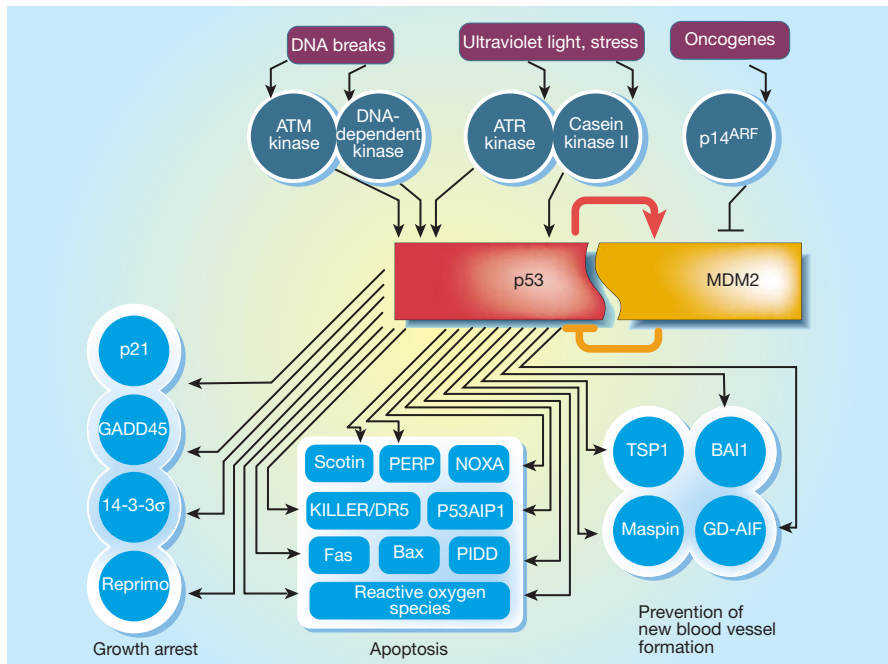


Figure 3 The p53 network. Activation of the network (by stresses such as DNA damage, ultraviolet light and oncogenes) stimulates enzymatic activities that modify p53 and its negative regulator, MDM2. This results in increased levels of activated p53 protein. The expression of several target genes is then activated by binding of the activated p53 to their regulatory regions. These genes are involved in processes that slow down the development of tumours. For example, some genes inhibit cell-cycle progression or the development of blood vessels to feed a growing tumour; others increase cell death (apoptosis). A negative feedback loop between MDM2 and p53 restrains this network. Many other components of this network, not shown here, have been identified. Similarly, p53 activation results in a variety of other effects, including the maintenance of genetic stability, induction of cellular differentiation, and production of extracellular matrix, cytoskeleton and secreted proteins. The components of the network, and its inputs and outputs, vary according to cell type. p53 is a highly connected 'node' in this network. It is therefore unsurprising that the loss of p53 function is so damaging, and that such loss occurs in nearly all human cancers.

The second pathway for activating p53 involves the expression of oncogenes in the absence of DNA damage^{4,5}. These oncogenes stimulate the transcription of the p14^{ARF} gene or stabilization of the p14^{ARF} protein, which then binds to MDM2 and inhibits its activity. There is also a spatial element to the regulation of MDM2 by p14^{ARF}. The p14^{ARF} protein is located within the nucleolus — a subcompartment within the nucleus. In some situations, p14^{ARF} appears to sequester MDM2 into this subcompartment. This keeps MDM2 away from p53, which remains outside the nucleolus (but within the nucleus) where it can activate the transcription of its target genes (see next section). Both MDM2 and p53 proteins also contain nuclear-import and nuclear-export signals — address labels that enable them to be directed into the nucleus and out again¹⁵. This offers yet another avenue for regulation. Indeed, p53 has been shown to reside outside the nucleus in some tumours¹⁵ (Fig. 1).

The study of viral oncogenes has also shown that interconnected signalling pathways control the activity of p53. Some DNA viruses — such as simian virus 40, human papilloma virus and adenoviruses — encourage the cells they infect to become

cancerous (Fig. 2). After infecting the cells of their hosts, these viruses produce proteins that bind to and inhibit another tumour suppressor, the retinoblastoma protein¹⁶, as well as proteins that inactivate p53.

These results have obvious implications for tumour development. But they also suggest that even a complete characterization of the genome (all the genes in an organism), the transcriptome (the genes that are actually expressed as mRNA at a given time) and the proteome (the proteins that are produced from the expressed genes) would not provide a very accurate portrait of the state of the p53 protein in any cell. The condition of this protein cannot be accurately predicted from just its sequence, as it is extensively 'decorated' by different chemical groups, rather as a Christmas tree is decorated by lights and tinsel.

As well as the covalent modifications described above, numerous proteins bind to p53 and may modify its stability as well as its ability to activate transcription⁸. Moreover, in a damaged or stressed cell there is not a single, monolithic p53 species but rather a variety, each modified in a specific fashion. And any detailed characterization of p53 must also include the fourth dimension — time. The state of p53 can change rapidly as

cells adapt to the network-initiating stimulus and respond to the numerous feedback and feedforward systems that are thereby set in motion.

What happens next?

Many biochemical functions have been ascribed to activated p53, but the best documented is its ability to bind to specific sequences in DNA and activate the transcription of adjacent genes¹⁷. The regions of p53 responsible for binding to specific sequences and activating transcription have been precisely defined. Virtually all naturally occurring mutations in the p53 gene reduce the ability of the encoded p53 protein to activate transcription, supporting the idea that this activity is critical to p53's role as a tumour suppressor.

Several dozen genes that are controlled directly by p53 have been identified¹⁷, and they fall broadly into four categories.

Cell-cycle inhibition. One of the first effects of p53 expression, in nearly all mammalian cell types, is a block in the cell-division cycle. The p53 protein directly stimulates the expression of p21^{WAF1/CIP1}, an inhibitor of cyclin-dependent kinases (CDKs). CDKs are key regulators of the cell cycle, working together with their partners — cyclin proteins — to ensure that, for example, DNA replication ('S phase') follows smoothly from the cellular resting phase known as G1.

Through its negative effects on various CDKs, p21^{WAF1/CIP1} inhibits both the G1-to-S and the G2-to-mitosis transitions. Other genes, the newest of which is Reprimo, can also arrest cells in G2 phase¹⁸. In epithelial cells — those that line organs such as the intestine and bladder — p53 also stimulates the expression of protein 14-3-3σ, which sequesters cyclin B1-CDK1 complexes outside the nucleus and thereby helps to maintain a G2 block^{19,20}. Interestingly, the inhibition of 14-3-3σ can, in a single step, make primary human epithelial cells grow indefinitely in culture²¹. This immortality may be a key feature distinguishing tumour cells from normal cells.

Apoptosis. Some cells in which p53 is activated undergo programmed death²². There are several potential mediators of p53-induced apoptosis¹⁷. The Bax protein — the prototype of this class of mediator — is an apoptosis-inducing member of the Bcl-2 protein family. Transcription of the Bax gene in some human cells is directly activated by p53-binding sites in the regulatory region of the gene²³. However, there is no analogous p53-binding site in the regulatory region of the murine Bax gene²⁴. More recently, the NOXA and P53AIP1 genes have been discovered to be directly activated by p53 (refs 25, 26). Like Bax, the NOXA and P53AIP1 proteins are located in mitochondria — the cellular powerhouses. When overexpressed, these proteins induce apoptosis.

Other potential mediators of p53-induced apoptosis include proteins with similarities to the classic 'death-signal' receptors, the TNF (tumour necrosis factor) receptor and Fas. The most recently discovered of these proteins is called PIDD²⁷. Finally, p53 may cause death by directly stimulating mitochondria to produce an excess of highly toxic reactive oxygen species.

Genetic stability. Not all genes that limit tumour development control cell birth or death directly. For example, repair genes involved in correcting certain types of error in DNA lead only indirectly to tumour development when inactivated (Box 1). This is because inactivation of these genes leads to genetic instability — an accumulation of errors in all genes, including those that control cell growth. The p53 protein may be important in maintaining genetic stability^{28,29}. The mechanisms are not clear, but they may involve the induction of genes that regulate 'nucleotide-excision' repair of DNA, chromosomal recombination and chromosome segregation^{28,29}. Further evidence for a role for p53 in DNA repair comes from the induction of a specific 'ribonucleotide reductase' gene by p53 after DNA damage³⁰. Such genes maintain cellular control over the responses to DNA errors in a wide range of organisms³¹.

Inhibition of blood-vessel formation. To reach dangerous sizes, tumours must encourage the growth of new nutrient-bringing blood vessels in their vicinity. The normal p53 protein stimulates the expression of genes that prevent this process^{17,32}. Cells in which p53 is inactivated by mutation would therefore be more likely to recruit new blood vessels, providing a critical growth advantage at a late point in tumour development. This stage is the time when most natural p53 mutations occur. Studies of other tumour suppressors support the idea that preventing the formation of new blood vessels can be an important component of the activity of a tumour suppressor³³.

As well as the genes that are directly activated by p53, there are many that are repressed, although the mechanisms involved are unclear^{34,35}. Why does p53 regulate the expression of so many genes — are most of them 'artefacts', which coincidentally have p53-binding sites, but do not have an important role in the p53 network?

Some p53-responsive genes are, in fact, activated only by artificially high levels of added p53, so the possibility of artefacts must be considered for any individual candidate gene. But alternative theories are more attractive and instructive, for the following reasons. First, even the most basic features of the cellular response to DNA damage are not the same in different species and vary even in different cells of the same organism^{34,35}. For example, high levels of normal p53 cause some human cells to undergo apoptosis, whereas others simply undergo prolonged

cell-cycle arrest. Therefore, one should expect variations in the expression of p53 target genes. Second, the genes most likely to be mutated in cancers, such as p53, are those that serve as nodal points for the integration of a large number of different signals. On this basis, one would expect there to be numerous downstream mediators of such genes, as explained below.

The p53 network

How can the vast number of activating signals, covalent and non-covalent modifications, and downstream regulators of p53 be put into context? One way to understand the p53 network is to compare it to the Internet. The cell, like the Internet, appears to be a 'scale-free network': a small subset of proteins are highly connected (linked) and control the activity of a large number of other proteins, whereas most proteins interact with only a few others³⁶. The proteins in this network serve as the 'nodes', and the most highly connected nodes are 'hubs'. In such a network, performance is almost unchanged by random removal of nodes. But such systems contain an Achilles' heel: "the most effective way of destroying a network is to attack its most connected nodes"³⁷. It is clear that p53 is one of the most highly connected nodes in the cell (Fig. 3, previous page), and that an attack on p53 (by mutation) will disrupt basic cellular functions, particularly the responses to DNA damage and tumour-predisposing stresses.

This theoretical framework has several implications. We should not be surprised that the inactivation of less connected nodes does not necessarily have draconian effects on the cell, and does not recapitulate all the effects of p53 inactivation. Instead, any outcome that mimics a part of what happens after p53 inactivation should be considered positive evidence of a link. For example, inactivation of the p21^{WAF1/CIP1} or Bax genes does not have exactly the same effects as inactivation of p53. Yet inactivation of each of these genes has tumour-enhancing effects in some cells under specific conditions^{23,38}.

One should also expect that combined attacks on many nodes linked to p53 should have progressively more severe effects that more and more closely resemble an attack on p53. This is exactly what has been observed^{39,40}. And finally, one should be especially cautious about experiments in which p53-linked nodes are overexpressed rather than disrupted. Such overexpressed gene products might interact with many other nodes in an abnormal and unregulated manner. Making sense of such experiments is much more difficult and error-prone than interpreting gene-disruption experiments. The same principles probably apply to most other tumour suppressors, which also function as highly connected nodes that respond to diverse influences within cell-type-specific networks.

An appreciation of the existence and

complexity of cellular networks should enable more rational design and interpretation of experiments in the future, and should allow more realistic approaches to treatment. After all, the most important question in p53 research is: how do we attack a cellular network that is already compromised by inactivation of one of its most highly connected nodes? New work⁴¹ suggests possible tactics for such an attack — and ways to dramatically affect the management of a diverse array of cancers.

Bert Vogelstein is at The Howard Hughes Medical Institute and Johns Hopkins Oncology Center, Baltimore, Maryland 21231, USA.

e-mail: vogelbe@welch.jhu.edu

David Lane is in the Department of Surgery and Molecular Oncology, Ninewells Hospital, University of Dundee, Dundee DD1 5EH, UK.

e-mail d.p.lane@dundee.ac.uk

Arnold J. Levine is in the Laboratory of Cancer Biology, Genetics, and Molecular Biophysics, Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

e-mail: alevine@rockvax.rockefeller.edu

- Hollstein, M. *et al.* *Mutat. Res.* **431**, 199–209 (1999).
- Hussain, S. P. & Harris, C. C. *Mutat. Res.* **428**, 23–32 (1999).
- Carr, A. M. *Science* **287**, 1765–1766 (2000).
- Sherr, C. J. & Weber, J. D. *Curr. Opin. Genet. Dev.* **10**, 94–99 (2000).
- Lowe, S. W. & Lin, A. W. *Carcinogenesis* **21**, 485–495 (2000).
- Meek, D. W. *Oncogene* **18**, 7666–7675 (1999).
- Momand, J., Wu, H. H. & Dasgupta, G. *Gene* **242**, 15–29 (2000).
- Prives, C. & Hall, P. A. *J. Pathol.* **187**, 112–126 (1999).
- Giaccia, A. J. & Kastan, M. B. *Genes Dev.* **12**, 2973–2983 (1998).
- Selivanova, G., Kawasaki, T., Ryabchenko, L. & Wiman, K. G. *Semin. Cancer Biol.* **8**, 369–378 (1998).
- Buschmann, T., Fuchs, S. Y., Lee, C. G., Pan, Z. Q. & Ronai, Z. *Cell* **101**, 753–762 (2000).
- Bar-Or, R. L. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 11250–11255 (2000).
- Lozano, G. & Liu, G. *Semin. Cancer Biol.* **8**, 337–344 (1998).
- Bell, D. W. *et al.* *Science* **286**, 2528–2531 (1999).
- Voussden, H. & Van de Woude, G. F. *Nature Cell Biol.* **2**, E178–E180 (2000).
- Levine, A. J. *Cell* **88**, 323–331 (1997).
- El-Deiry, W. S. *Semin. Cancer Biol.* **8**, 345–357 (1998).
- Ohki, R. *et al.* *J. Biol. Chem.* **275**, 22627–22630 (2000).
- Chan, T. A. *et al.* *Nature* **401**, 616–620 (1999).
- Laronga, C. *et al.* *J. Biol. Chem.* **275**, 23106–23112 (2000).
- Dellambra, E. *et al.* *J. Cell Biol.* **149**, 1117–1130 (2000).
- Gottlieb, T. M. & Oren, M. *Semin. Cancer Biol.* **8**, 359–368 (1998).
- Reed, J. C. *J. Clin. Oncol.* **17**, 2941–2953 (1999).
- Schmidt, T. *et al.* *Cell Death Differ.* **6**, 873–882 (1999).
- Oda, E. *et al.* *Science* **288**, 1053–1058 (2000).
- Oda, K. *et al.* *Cell* **102**, 849–862 (2000).
- Lin, Y., Ma, W. & Benichou, S. *Nature Genet.* **26**, 122–127 (2000).
- Tlsty, T. D. *Curr. Top. Microbiol. Immunol.* **221**, 37–46 (1997).
- Wahl, G. M. *et al.* *Cancer Surv.* **29**, 183–219 (1997).
- Tanaka, H. *et al.* *Nature* **404**, 42–49 (2000).
- Lozano, G. & Elledge, S. J. *Nature* **404**, 24–25 (2000).
- Hendrix, M. J. *Nature Med.* **6**, 374–376 (2000).
- Schwarte-Waldhoff, I. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 9624–9629 (2000).
- Yu, J. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 14517–14522 (1999).
- Zhao, R. *et al.* *Genes Dev.* **14**, 981–993 (2000).
- Albert, R. *et al.* *Nature* **406**, 378–382 (2000).
- Tu, Y. *Nature* **406**, 353–354 (2000).
- Dotto, G. P. *Crit. Rev. Oral Biol. Med.* **10**, 442–457 (1999).
- Jones, J. M. *et al.* *Cell Growth Differ.* **10**, 213–222 (1999).
- Franklin, D. S. *et al.* *Mol. Cell Biol.* **20**, 6147–6158 (2000).
- Lamont, J. P. *et al.* *Ann. Surg. Oncol.* **7**, 588–592 (2000).

Related websites

♦ <http://www.cancergenetics.org/p53.htm>

♦ <http://www.iarc.fr/p53/Index.html>

Tracing the geographical origin of cocaine

Cocaine carries a chemical fingerprint from the region where the coca was grown.

Here we show that cocaine originating from different geographic regions in South America can be identified by its isotope-ratio signature. The distinct carbon ($\delta^{13}\text{C}$) and nitrogen ($\delta^{15}\text{N}$) isotope-ratio combinations allow the country of origin to be determined for the principal coca-growing regions along the Andean Ridge. By combining this information with detectable differences in the patterns of the trace alkaloids truxilline and trimethoxycocaine, we correctly identified the source of 96% of 200 cocaine samples.

Cocaine is the most widely used narcotic drug, making the determination of the geographic origin of illicit cocaine the focus of intense investigation by the forensic community^{1,2}. Previous studies have concentrated on detecting the trace residues present in cocaine or the trace alkaloids that are extracted with it; these have met with limited success, although they have been valuable in identifying the processing methods peculiar to different regions. However, refined cocaine base is often transported from one country to another for its final conversion to cocaine hydrochloride, making the source harder to identify.

Stable-isotope ratios (δ) have been used as indicators of the geographical source of a wide variety of biological and non-biological materials, from migrating butterflies³ to emeralds⁴. The natural isotopic variation in organic matter in plants can be applied to predict the variation in $\delta^2\text{H}$, $\delta^{13}\text{C}$, $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ values along ecological gradients⁵⁻⁸.

There are large environmental differences in the coca-growing regions of South America. We investigated the variation in $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ of coca leaves and of the extracted cocaine from plants from different countries along the Andean Ridge, collecting 200 coca leaf sets from throughout each of the five primary growing regions: the Chapare Valley of Bolivia, the Huallaga and Ucayali Valleys and Apurimac Valley of Peru, and the Putumayo–Caqueta and Guaviare regions of Colombia (Fig. 1). Subsets of the same coca leaves were then extracted for determination of total alkaloids⁹, and trimethoxycocaine and truxilline

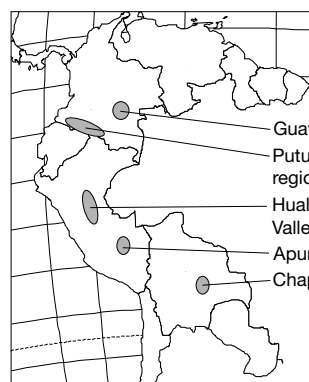
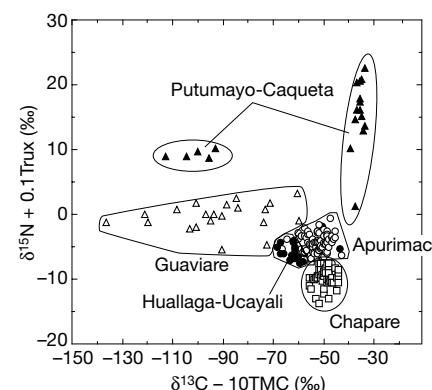


Figure 1 Identification of geographic regions in South America where coca is commonly grown. Left, regions producing illicit cocaine; right, identification of cocaine-growing regions based on a combined model derived from carbon- and nitrogen-isotope ratios as well as abundance of minor alkaloid components. Squares, Bolivia; triangles, Colombia; and circles, Peru. Regions within a country are distinguished by black and white symbols. Trux, truxilline; TMC, trimethoxycocaine. Isotope ratios are expressed as $(R_{\text{sample}}/R_{\text{standard}} - 1) \times 1,000\text{‰}$, where R is the molar ratio of heavy-to-light stable isotope; standards for carbon and nitrogen are PDB and air, respectively.

content¹⁰. We purified cocaine base (to over 98% purity) from these extracts by column chromatography on alumina and measured $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ to a precision of greater than 0.1 part per thousand on corresponding sets of coca leaves and purified cocaine samples¹¹.

Over the entire geographical distribution, $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ values for coca leaves vary from -32.4‰ to -25.3‰ for $\delta^{13}\text{C}$, and from 0.1‰ to 13.0‰ for $\delta^{15}\text{N}$. Leaves from the Putumayo and Caqueta regions of Colombia are distinguishable from each other by their $\delta^{13}\text{C}$ content, as are those from the Huallaga and Ucayali Valleys from the Apurimac Valley of Peru. Coca leaf from Bolivia has consistently less ^{15}N than material from Peru. There is more ^{15}N in coca leaves from Colombia and least in coca grown in the Chapare Valley of Bolivia.

To determine the probability that a sample could have originated from one of these three countries, we used bivariate mean and standard deviation parameters to estimate the frequency with which we could correctly identify the country of origin of these coca-leaf samples. We predicted the country of origin by assigning a sample to the region to which it was associated with the highest probability. Using this approach, we accu-



ately predicted the country of origin of 90% of the 200 coca-leaf samples.

The truxilline and trimethoxycocaine content of coca leaves contributed additional information, showing a clear demarcation between the Guaviare and Putumayo–Caqueta regions and locations in Peru and Bolivia (Table 1). Although it was not possible to differentiate Peruvian and Bolivian cocaine from their trace-alkaloid analyses, these samples were distinguishable by $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ analysis.

Combining alkaloid and isotopic-ratio analyses is a powerful means for determining the country of origin of cocaine from these different regions in South America (Fig. 1). Taking advantage of our observation that truxilline content and cocaine $\delta^{15}\text{N}$ are positively correlated and that cocaine $\delta^{13}\text{C}$ and trimethoxycocaine are negatively correlated, a combination plot reveals that the cocaine samples from different regions cluster into tighter groups based on their country of origin. Using our analytical approach, we were able accurately to predict the country of origin of 96% of the 200 cocaine samples.

The environmental basis for this regional variation in coca-leaf and cocaine stable isotopes may be due to differences in soils affecting $\delta^{15}\text{N}$, and in the humidity and length of the wet season affecting $\delta^{13}\text{C}$. It has been proposed⁸ that nutrient cycles in tropical forest ecosystems (typical of coca-growing regions) are more open than in drier, more temperate regions, and the range of coca-leaf $\delta^{15}\text{N}$ values are consistent with this idea. $\delta^{13}\text{C}$ differences are narrower, but are consistent with patterns predicted by $\delta^{13}\text{C}$ theory for plants^{5,6}.

Table 1 Differences in trace alkaloid amounts among coca populations

Region	Truxillines	Trimethoxycocaines	Ratio
Chapare Valley, Bolivia	2.79 ± 0.41	0.16 ± 0.03	17.2
Huallaga/Ucayali Valleys, Peru	3.80 ± 0.76	0.19 ± 0.04	20.0
Apurimac Valley, Peru	4.11 ± 1.59	0.27 ± 0.06	15.2
Guaviare region, Colombia	4.94 ± 2.20	0.61 ± 0.06	8.1
Putumayo–Caqueta region, Colombia	14.66 ± 4.29	0.20 ± 0.29	7,330

Means and standard deviations of truxilline and trimethoxycocaine content of coca leaves, together with the ratio of truxillines to trimethoxycocaines growing in different regions of South America are shown. Data are given on a w/w% basis using structurally related internal standards.

Tracing the country of origin of cocaine is now feasible through automated, routine analysis of both stable isotopes and trace alkaloids, opening up strategic options for identifying source regions and trafficking routes. We have shown how ecological and isotopic-fractionation principles used to predict isotopic-ratio patterns associated with plants from different ecosystems can also be applied to determine the distribution of an illegal drug, as well as to identify new coca-producing regions as they develop.

James R. Ehleringer*, John F. Casale†, Michael J. Lott*, Valerie L. Ford†

*Stable Isotope Ratio Facility for Environmental Research, Department of Biology, University of Utah, 257 S 1400 E, Salt Lake City, Utah 84112-0840, USA

e-mail: ehleringer@biology.utah.edu

†Special Testing and Research Laboratory, Drug Enforcement Administration, US Department of Justice, 7704 Old Springhouse Road, McLean, Virginia 22102-3494, USA

- Moore, J. M. & Casale, J. F. *Forensic Sci. Rev.* **10**, 13–46 (1998).
- Moore, J. M. & Casale, J. F. *J. Chromatogr. A* **674**, 165–205 (1994).
- Wassenaar, L. I. & Hobson, K. A. *Proc. Natl Acad. Sci. USA* **95**, 15436–15439 (1998).
- Giulani, G. et al. *Science* **287**, 631–633 (2000).
- Farquhar, G. D., Ehleringer, J. R. & Hubick, K. T. *Annu. Rev. Plant Physiol. Mol. Biol.* **40**, 503–537 (1989).
- Ehleringer, J. R., Hall, A. E. & Farquhar, G. D. (eds) *Stable Isotopes and Plant Carbon/Water Relations* (Academic, San Diego, 1993).
- Roden, J. S., Lin, G. & Ehleringer, J. R. *Geochim. Cosmochim. Acta* **64**, 21–35 (2000).
- Martinelli, L. A. et al. *Biogeochemistry* **46**, 45–65 (1999).
- Moore, J. M., Casale, J. F., Klein, R. F., Cooper, D. A. & Lydon, J. *J. Chromatogr. A* **659**, 163–175 (1994).
- Moore, J. M., Casale, J. F. & Cooper, D. A. *J. Chromatogr. A* **756**, 193–201 (1996).
- Ehleringer, J. R., Cooper, D. A., Lott, M. J. & Cook, C. S. *Forensic Sci. Intl* **106**, 27–35 (1999).

Archaeology

Detecting milk proteins in ancient pots

Deciding whether to farm cattle for milk or beef was just as complex in the past as it is today. Compared with meat production, dairying is a high-input, high-output, high-risk operation indicative of an intensive, sophisticated economy, but this practice is notoriously difficult to demonstrate in the archaeological record¹. Here we provide evidence for the presence of milk proteins preserved in prehistoric vessels, which to our knowledge have not been detected before. This finding resolves the controversy that has surrounded dairying on the Scottish Atlantic coast during the Iron Age^{2–5} and indicates that farming by the early inhabitants of this harsh, marginal environment was surprisingly well developed.

The analysis of sorbed lipid residues in prehistoric ceramics has provided a powerful new indicator of how vessels were

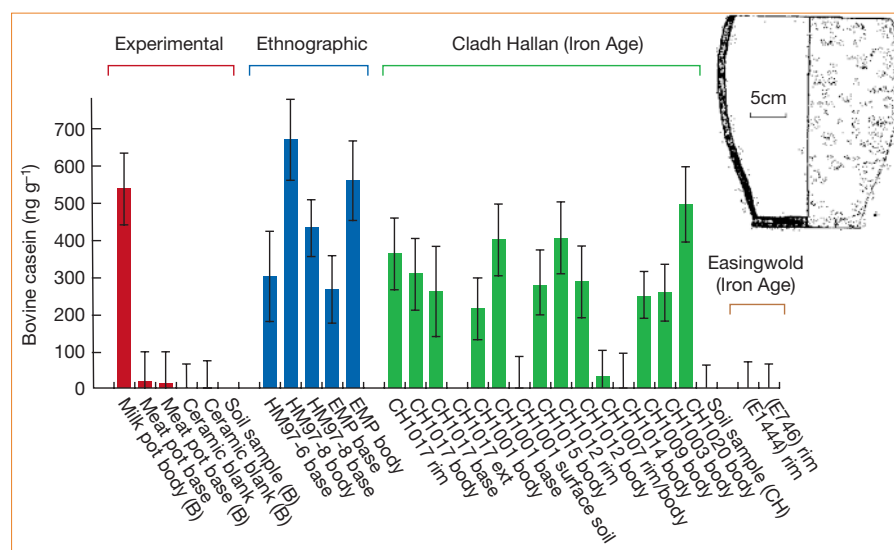


Figure 1 Amounts of bovine α -casein present in samples of pottery and soil, as determined by duplicate assay using digestion-and-capture immunoassay with a monoclonal antibody raised against this protein. Error bars, one standard deviation. The assay is specific only for cows' milk and is able to detect as little as 100 ng protein per g of ceramic (0.1 p.p.m.). Experimental coarseware pots (ceramic 'blank') were used to boil either milk (milk pot) or beef (meat pot) repeatedly and were buried for 1 year in upland soil. Ethnographic pots were obtained from Pakistan (HM) and India (EMP); each had been recently used to prepare dairy products. Cladh Hallan (CH) vessels (inset) were collected from a single site (fill of house 112, South Uist, Outer Hebrides). Domestic cooking pots from Easingwold, North Yorkshire (E), contained large amounts of well-preserved animal fats.

used^{6–8}. Although proteins are more diagnostic of specific foodstuffs than lipids, they are difficult to extract from archaeological ceramics⁹. We have developed an immunological detection method, the digestion-and-capture immunoassay (DACIA)¹⁰, which overcomes this difficulty by dissolving the ceramic then capturing the liberated proteins for immunodetection.

We obtained sherds from nine coarseware cooking vessels, dated to the middle of the first millennium BC, from the fill of an Early Iron Age house at Cladh Hallan, South Uist, in the Outer Hebrides, and analysed them by DACIA. Extracts were tested using a monoclonal antibody raised against heat-degraded and dephosphorylated bovine α -casein (about 1.4% w/v milk), which was specific for bovine milk.

Immunological analysis of archaeological materials has been criticized for the lack of negative controls¹¹, so we included an extensive array of reference samples (Fig. 1). Seven of nine of the interior sections of sherds recovered from Cladh Hallan tested positive for casein and the amounts were comparable to those found on experimentally buried milk sherds (Fig. 1). DACIA analysis failed to detect the presence of bovine α -casein in the associated sediment or exterior surfaces of the samples.

The large number of neonatal cattle remains found at this site (42% of individuals) has been attributed to the deliberate culling of young calves in order to preserve fodder in an adverse environment^{2,3} or to sustain a high-input dairying economy^{4,5}. The presence of bovine α -casein on a substantial number of sherds (Fig. 1) lends

support to the latter interpretation. Our successful characterization of protein residues after 2,500 years demonstrates the potential of DACIA as a high-resolution technique for determining how archaeological ceramics were used.

Oliver Craig*, Jacqui Mulville†, Mike Parker Pearson‡, Robert Sokol§, Keith Gelsthorpe§, Rebecca Stacey||, Matthew Collins*

*Fossil Fuels and Environmental Geochemistry, NRG, Drummond Building, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne NE1 7RU, UK

e-mail: o.e.craig@ncl.ac.uk

†English Heritage, Oxford University Museum, Parks Road, Oxford OX1 3PW, UK

‡Department of Archaeology and Prehistory, University of Sheffield, Northgate House, West Street, Sheffield S1 4ET, UK

§Regional Blood Transfusion Centre, Longley Lane, Sheffield S5 7JN, UK

||Department of Archaeological Science, University of Bradford, Bradford BD7 1DP, UK

- Halstead, P. *Anthropozoologica* **27**, 3–20 (1998).
- McCormick, F. in *Life on the Edge: Human Settlement and Marginality* (eds Mills, C. M. & Coles, G.) 49–53 (Oxbow, Oxford, 1998).
- Gilmour, S. & Cook, M. *Antiquity* **72**, 327–337 (1998).
- Parker-Pearson, M., Sharples, N. & Mulville, J. *Antiquity* **70**, 57–67 (1996).
- Parker-Pearson, M., Sharples, N. & Mulville, J. *Antiquity* **73**, 149–152 (1999).
- Rottländer, R. C. A. in *Proc. 24th Int. Symp. Archaeometry* 403–405 (Smithsonian Institution, Washington DC, 1986).
- Evershed, R. P. *World Archaeol.* **25**, 74–93 (1993).
- Dudd, S. N. & Evershed, R. P. *Science* **282**, 1478–1481 (1998).
- Evershed, R. P. & Tuross, N. *J. Archaeol. Sci.* **23**, 429–436 (1996).
- Craig, O. E. & Collins, M. J. *J. Immunol. Meth.* **236**, 89–97 (2000).
- Tuross, N., Barnes, I. & Potts, R. *J. Archaeol. Sci.* **23**, 289–296 (1996).

Feedback control of intercellular signalling in development

Matthew Freeman

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

The intercellular communication that regulates cell fate during animal development must be precisely controlled to avoid dangerous errors. How is this achieved? Recent work has highlighted the importance of positive and negative feedback loops in the dynamic regulation of developmental signalling. These feedback interactions can impart precision, robustness and versatility to intercellular signals. Feedback failure can cause disease.

Most animal cells develop according to cues in their environment that have been produced by other cells. In the past twenty years many of these signalling molecules and their transduction pathways have been identified, and this has led to the discovery that the same signalling pathways appear in several developmental contexts. For example, a signal that in one instance will cause a cell to differentiate terminally will elsewhere lead another cell type to undergo mitosis and in a third context will trigger cell death. In general, the outcome of the signalling event is not determined by the signal itself but by the developmental state of the cell receiving it. This developmental state consists in the various cellular targets (such as transcription factors and cytoskeletal proteins) that are primed to respond to the signal; whether they are primed depends on the history of the cell.

Obviously, developmental signalling events must be precisely regulated. A signal that is produced in the wrong time or place will lead to inappropriate developmental responses, which can be dangerous—cells must be protected against this. Also, desirable signals must be robust enough to ensure that cells receive them at high enough levels to respond. Just as important is versatility. Not only is there a wide range of different cell types and tissue environments in which these signals must operate, but they must also function with different spatial and kinetic properties. These three main properties of intercellular signalling in development—precision, robustness and versatility—are stringent requirements and errors are serious. How is this control achieved?

Feedback loops can account for aspects of all these properties. Feedback can be defined as the ability of a system to adjust its output in response to monitoring itself. More than 2,000 years ago, complex systems were designed to incorporate feedback principles (an accurate water clock from the third century BC in Alexandria is the earliest known example of feedback in engineering), although it was only formalized as a theoretical branch of science and engineering by Wiener in the 1940s¹. Two prominent areas of feedback in biology that have been studied historically have been general growth theory², in which cell growth was postulated to be regulated by inhibitory factors produced by cells themselves, and theoretical models of pattern formation^{3–5}.

More recently, as developmental biology has acquired the molecular tools that allow mechanistic insight, it has become clear that the principles of feedback—both positive and negative loops—are indeed used to produce the signalling properties necessary in animal development. Negative feedback occurs when, for example, a signal induces the expression of its own inhibitor; it serves to dampen and/or limit signalling. Positive feedback occurs when a signal induces more of itself, or of another molecule that amplifies the initial signal, and this serves to stabilize, amplify or prolong signalling. The

widespread use of feedback and the variety of its consequences make it an important principle of regulating signals in development.

Here I describe some examples of developmental feedback control, in order to illustrate the strategies in which it participates. The focus is particularly on experimental demonstration of feedback in intercellular signalling, rather than the extensive literature on theoretical modelling of feedback, especially in the context of networks of transcription factors⁶. The issues that will be covered include temporal and spatial control of signalling, and the unification of these into complex patterning and growth control. The relevance of feedback in cancer—where normal developmental controls are disrupted—is also discussed.

Temporal control of signalling

Perhaps the most obvious use of negative feedback is to limit the duration of a signal. In its simplest form, a signal induces its own negative regulator so that when a threshold has been reached, the signal ceases. An example of this is the control of cytokine signalling through the JAK/STAT signalling pathway (Fig. 1). JAKs are soluble tyrosine kinases that bind to cytokine receptors and transduce signals by the STAT proteins: transcriptional activators with SH2 domains that bind to the phosphotyrosine on activated JAKs⁷. Since 1997, a growing family of cytokine-inducible proteins (variously termed SOCS, SSI, JAB and CIS^{8–10}) that inhibit cytokine signalling has been identified^{11,12}. These proteins participate in negative feedback loops. The physiological significance of the negative feedback is exemplified by SOCS1, which is induced by and inhibits interferon- γ signalling; when the gene for SOCS1 is knocked out, mice die as neonates with defects associated with excess interferon- γ signalling^{13,14}.

Other transgenic and knockout experiments have shown that SOCS3 is necessary for the control of erythropoiesis¹⁵, probably through a negative feedback loop controlling the response to the cytokine erythropoietin. SOCS3 also participates in negative feedback control of metabolic factors including leptin and growth hormone^{16,17}. Although the mechanisms of SOCS inhibition are not yet fully understood, there are indicators: SOCS1 acts as a pseudosubstrate inhibitor of JAK activity, and the SOCS domain binds to elongins, thereby targeting proteins for proteasomal degradation.

Less obvious than negative feedback control of signalling kinetics is the use of positive feedback loops to prolong signalling. A good example of this occurs in the developing *Drosophila* egg, where epidermal growth factor (EGF)-receptor signalling is crucial for, among other things, specification of dorsal follicle cells¹⁸. The signal is initiated by the transforming growth factor (TGF)- α -like ligand Gurken, produced by the oocyte, and received by the EGF receptor in the overlying somatic follicle cells (Fig. 2). However, this signal

must be short-lived because soon after it is initiated, an impermeable layer called the vitelline membrane develops between the oocyte and follicle cells. This physical barrier to signalling is overcome when Gurken induces the follicle cells to start producing an autocrine signal—another TGF- α ligand, Spitz—to prolong (and amplify) the initial EGF receptor activity¹⁹.

Spatial control by feedback

Spatial regulation of developmental signalling is also regulated by feedback. A good example is the expression of homeotic genes that control development in all animals. Their tissue-specific expression must be maintained throughout development—long after the initial localized signals that establish them have died away²⁰. Positive feedback loops assist this maintenance in a number of cases. This 'autoregulation' of homeotic genes occurs because they activate their own transcription, causing a stabilization and maintenance of their initial expression patterns. In principle, this positive feedback loop could be direct (the homeotic genes encode transcription factors) or indirect, involving intermediate steps. Both types of feedback have been identified: strikingly, not only do homeotic proteins directly activate their own transcription, but they also autoregulate through positive feedback loops that involve intercellular communication.

The most studied example of this is the positive feedback loop that maintains the expression of the *Ubx* gene in parasegment 7 of the *Drosophila* gut^{21,22} (Fig. 3). Here, *Ubx* controls the expression of the bone morphogenetic protein (BMP)-2/4-like protein Dpp, which in turn causes cells in the adjacent parasegment 8 to express the Wnt protein, Wingless; Wingless then signals back to the parasegment 7 cells. These respond to the combination of Wingless and Dpp signals by activating *Ubx* transcription, thereby closing the feedback loop. This system appears to play an important part in the spatial fine-tuning of *Ubx* expression, as well as its more obvious function in maintenance and stability. It acts to integrate the consequences of two different signalling molecules from different sources (note that Wingless expression in parasegment 8 is also under other spatial control), and thereby ensures that *Ubx* expression is restricted to the correct region of the gut²³. The autoregulation of other *Drosophila* and mammalian homeodomain proteins has also been found to involve positive feedback signalling

loops^{24–27}, suggesting that this indirect mechanism is a well-conserved principle.

The *Drosophila* EGF receptor antagonist Argos^{28,29} also participates in spatial feedback control, this time in a negative feedback loop. Argos expression is induced by EGF receptor signalling³⁰, establishing a negative feedback loop—one of the earliest whose developmental significance was understood. In this case, the antagonistic ligand Argos acts at a longer range than the agonist, the TGF- α -like Spitz. This sets up a system of 'remote inhibition' during ommatidial development in the fly eye: cells closest to the source of Spitz respond by producing Argos. These cells, however, are already irreversibly destined to differentiate and therefore unaffected by Argos. Argos, however, diffuses away from its source and prevents more remote cells, which are not yet committed, from responding to the low levels of Spitz to which they are exposed^{31,32}. This limits the effective range of EGF receptor signalling and thereby the number of cells that are able to be recruited to the developing eye.

In fact, the *Drosophila* EGF receptor pathway is regulated by a large number of feedback mechanisms (Fig. 4). The signalling inhibitors Argos, Sprouty and Kerkon1 are all induced by EGF receptor activity³³, as are the activating molecules Rhomboid-1 and Vein, at least in some contexts^{19,34,35}. This complexity of regulation raises the question of whether this is a case of regulatory overkill, or do these different feedback regulators have distinct functions? The evidence points to the latter—each of these feedback events has its own role and significance. For example, Argos is EGF-receptor-specific, secreted, and can act over many cell diameters^{28,29,36}. Sprouty is an intracellular inhibitor of Ras activation and can be induced by, and will inhibit signalling from, a variety of receptor tyrosine kinases³⁷.

The third known inhibitor, Kerkon1, is EGF-receptor-specific, but is a transmembrane protein, so its action is, unlike Argos, confined to the cell expressing it³⁸. Consistent with these theoretical differences, the three known feedback inhibitors of the EGF receptor have distinct phenotypes. Similarly, molecular and genetic analysis has demonstrated that the positive feedback mediated by Rhomboid, an activator of ligand production, and Vein, a neuregulin-like ligand for the EGF receptor, have different roles in controlling signalling.

Integration of feedback events in pattern formation

In reality, the distinction between temporal and spatial control of signalling is artificial. Cells often receive several spatially and temporally restricted signals and must integrate the information to respond appropriately. The most striking developmental examples of this involve pattern formation. How does communication

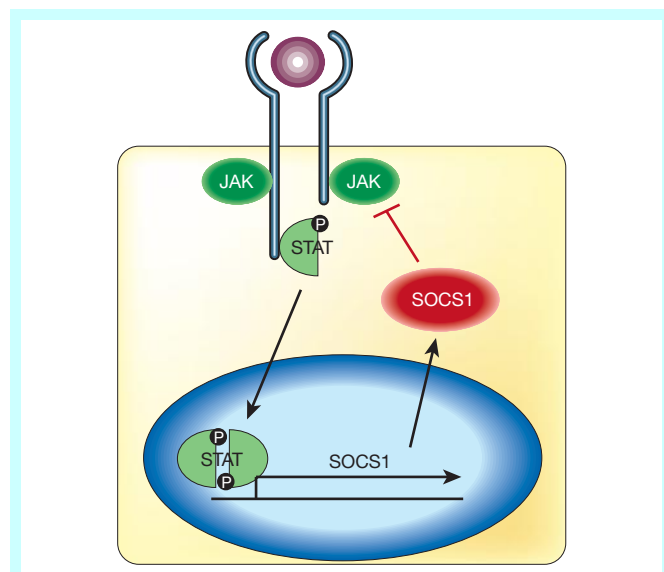


Figure 1 SOCS1 negative feedback loop. Ligand-bound cytokine receptors recruit JAKs, which in turn phosphorylate the transcriptional activator STAT proteins. SOCS1/CIS/SSI proteins are consequently expressed and inhibit JAK/STAT signalling.

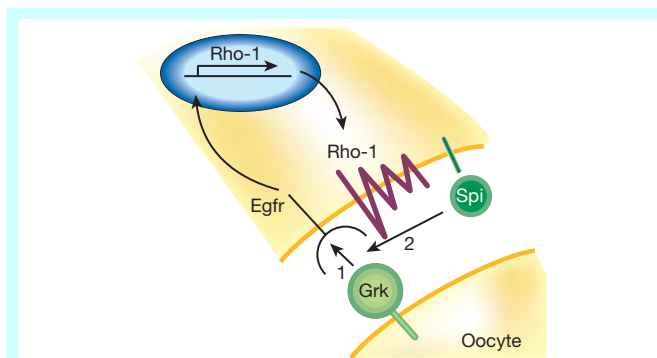


Figure 2 Autocrine amplification by positive feedback in the *Drosophila* oocyte. The TGF- α -like ligand, Gurken (Grk), expressed in the oocyte, activates the EGF receptor (Egfr) in the overlying follicle cells. This activates the expression of Rhomboid-1 (Rho-1), which in turn activates another EGF-receptor ligand, Spitz (Spi). Spitz then amplifies and prolongs EGF-receptor signalling in the follicle cells.

between cells lead to the patterning events needed to produce an animal? Again, positive and negative feedback control emerges as a significant factor. In some cases, the integration of successive feedback loops has been demonstrated so that extensive and complex patterning can now be better understood. Four recently discovered examples will be considered here, two from vertebrates and two from *Drosophila*.

Positive and negative feedback establish left–right asymmetry.

Recent progress has been made in understanding how the fundamental property of left–right asymmetry is established in vertebrate embryos. These leaps in understanding are reviewed elsewhere³⁹ so I will describe only the relevant details here. The principal factor appears to be the TGF- β family signalling molecule, Nodal⁴⁰, which is required throughout the left side of the embryo to activate left-specific genes. There are two key events in left–right specification: the initial symmetry breaking and its relay to all cells.

Symmetry breaking is not well understood in all animals but some excellent studies in the mouse embryo have shown that the first event is caused by specialized cilia in cells around the node⁴¹. These produce a vortical leftward flow of extracellular fluid within which, presumably, are growth factors that lead to the asymmetric expression of Nodal on the left side of the node. Nodal expression then needs to be relayed to the more distal cells of the lateral plate mesoderm (LPM), from which the left–right asymmetric organs differentiate.

This relay of *nodal* expression is now understood to involve a number of interlocking signalling relays that incorporate positive and negative feedback loops. Although there may be significant species differences in the relay mechanisms, the general and crucial function of Nodal in the left LPM seems to be global. Nodal, in complex with the extracellular cofactor EGF-CFC, signals to maintain its own expression, thereby participating in a positive feedback loop^{42–44}. But Nodal also induces the expression of its own inhibitor, Lefty-2/antivin, which competes for Nodal receptors and thus restricts its range of action⁴⁵. In chicks, another negative feedback loop involves the recently discovered Caronte, a TGF- β family factor that relays the localized Nodal expression to more distal cells^{46,47}. Thus, feedback control of signalling has a central role in the establishment of widespread Nodal expression, restricted to the left side of the embryo.

Positive feedback can coordinate distinct signals. The vertebrate limb provides a good example of another strategic role for feedback control. The growth and development of the outgrowing limb depends primarily on two organizing centres, the zone of polarizing

activity (ZPA) and the apical ectodermal ridge (AER) (Fig. 5)⁴⁸. The ZPA corresponds to a region of mesoderm at the posterior margin of the limb and it moves distally as the limb grows so that it is always on the posterior edge of the limb, close to the distal end. The AER is a ridge of ectoderm that corresponds to the distal tip of the limb. A positive feedback loop maintains the spatial relationship between these two crucial domains^{49,50}.

The secreted protein sonic hedgehog (Shh) is produced by the ZPA (it is still unclear how this is initiated) and this is transmitted, through a cascade of signalling molecules⁵¹, to the AER, where it maintains the expression of a fibroblast growth factor (this factor has been believed to be FGF-4, although recent evidence from FGF-4 mutants suggests that it may in fact involve other FGFs⁵²). The FGF secreted from the AER is then required for the upregulation and maintenance of Shh in the ZPA, thus producing a positive feedback loop between the ZPA and AER. This feedback loop coordinates two essential aspects of limb development: first, growth and proliferation of cells at the growing tip, which are controlled by FGFs; and second, the differentiated fate of cells, which is controlled by Shh from the ZPA.

Negative feedback can restrict ligand range. In the *Drosophila* wing, where much progress has been made in understanding the principles of patterning, a variation on the theme of negative feedback limits signalling by Hedgehog (Hh) to the anterior/posterior compartment boundary. Hh signalling induces the expression of the long-range morphogen Dpp (the fly homologue of BMP2/4), and the global patterning of the wing is therefore dependent on the restricted activity of Hh signalling^{53–55}. In this case, the key molecule in the negative feedback loop is Patched^{56,57}, a component of the Hh receptor that binds Hh but which does not have transducing activity (which is provided by another receptor component, Smoothened^{58,59}).

Patched is expressed at low levels in all cells in the anterior compartment, which are Hh-responsive, but is dramatically up-regulated in anterior cells near the A/P compartment boundary, when they receive Hh from their posterior compartment neighbours, the source of Hh (Fig. 6). In these cells near the anterior/posterior boundary, now expressing high levels of Patched, Hh is efficiently sequestered, thereby preventing it from diffusing more than a few cells away from the Hh source⁶⁰. Thus, Hh induces the expression of a molecule that restricts the range of Hh signalling.

The dependence of Patched expression on Hh signalling has been conserved throughout evolution^{61,62}, suggesting that this feedback

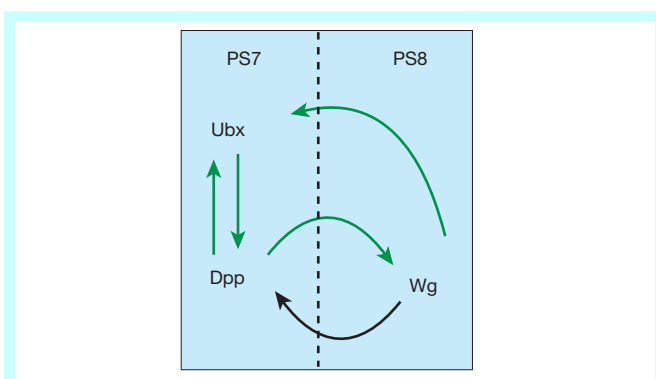


Figure 3 Ubx autoregulation by indirect positive feedback. The regulatory relationship between the three main players in the circuit is shown. The principal feedback loop is highlighted in green: Ubx in parasegment 7 activates the transcription of the signal Dpp; this leads to the expression of Wg in cells of the adjacent parasegment 8; the combination of Dpp and Wg is required to activate Ubx transcription. Ubx is repressed in parasegment 8 by the combined inhibitory action of another homeotic gene, Abd-A, and Wg (which is inhibitory at high concentrations).

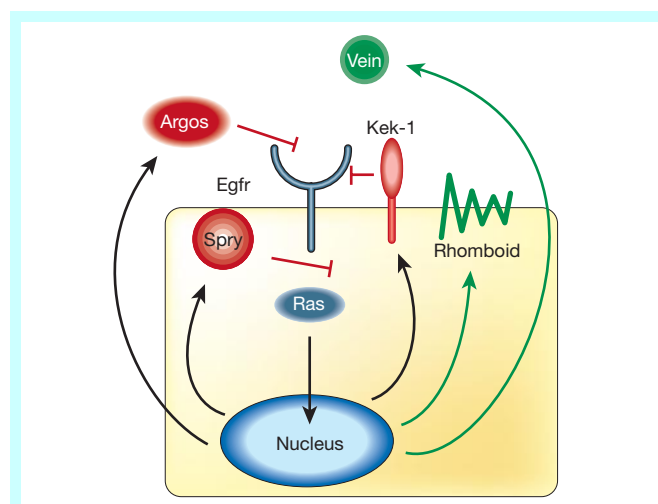


Figure 4 Multiple feedback loops regulate the *Drosophila* EGF receptor. There are three known feedback inhibitors (shown in red) of the fly EGF receptor (Egfr), which signals through Ras: Sprouty (Spry), Kekk-1 (Kek-1) and Argos. There are at least two components that act in positive feedback, shown in green: Rhomboid and Vein.

loop that limits the signalling range may also be conserved. As there are tissues in which Hh appears to act at long range, it is possible that the sensitivity of this induction of Patched can be tuned to allow different ranges of Hh signalling in distinct developmental contexts. Another negative feedback regulator of Hh signalling, discovered in vertebrates, is Hip, a non-signalling Hedgehog-binding protein, whose expression is regulated by Hh signalling⁶³.

Integrated feedback cycles can elaborate pattern. The examples described so far in this section have illustrated the importance of feedback control in patterning and how its strategic significance varies in different contexts. The final example of feedback in patterning—the specification of the dorsal/ventral axis and the dorsal appendages in the *Drosophila* egg—highlights how distinct feedback loops can themselves integrate to produce elaboration of pattern.

The anterior–dorsal region of the fly egg has two respiratory appendages that arise from either side of the dorsal midline (Fig. 7). Their differentiation and location is specified by EGF receptor signalling in the somatic follicle cells that surround the egg. As described above, a positive feedback loop occurs when EGF receptor signalling, which is initiated by the TGF- α -like ligand Gurken expressed in the oocyte, triggers the activation of a different TGF- α -like ligand, Spitz, in the follicle cells themselves. This autocrine amplification of an initially paracrine signal causes the peak levels of EGF receptor signalling in cells at the midline to reach a threshold required for the establishment of a negative feedback loop dependent on the expression of the secreted EGF receptor antagonist, Argos (Fig. 7). Argos inhibits signalling at the midline, thereby splitting the initial single peak of receptor signalling into twin peaks, and it is these that specify the position of the appendages¹⁹.

The positive and negative feedback loops are integrated because the amplification of the initial signal is required to induce Argos expression and thus trigger the secondary negative feedback loop. Another point that arises from this example is that the integrated feedback system controls not only the location of signalling but also its timing. The initial single peak at the midline is responsible for the determination of the dorsal/ventral axis, but the displaced twin peaks later determine the appendages. The integrated feedback system ensures that this temporal sequence—first one peak followed by two—occurs correctly.

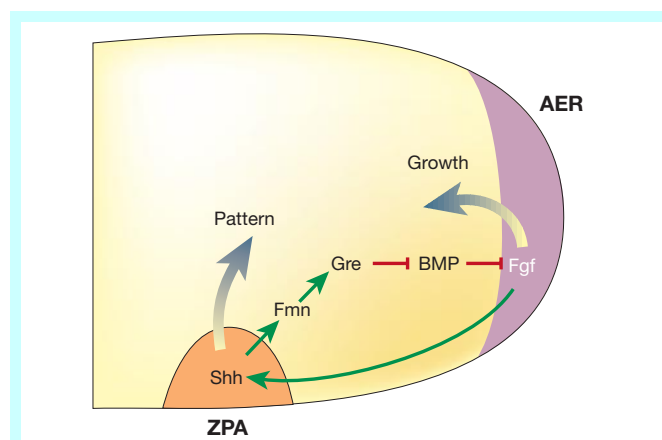


Figure 5 A positive feedback loop coordinates vertebrate limb development. The zone of polarizing activity (ZPA) and the apical ectodermal ridge (AER) are the two principal organizing centres of the limb, patterning the anterior/posterior axis and controlling proliferation at the growing tip, respectively. These two essential organizers of development are coordinated in the growing limb tip by a positive feedback loop that makes them mutually dependent. Sonic hedgehog (Shh) from the ZPA activates the expression of fibroblast growth factor (Fgf) in the AER, via a number of signalling components, Formin (Fmn), Gremlin (Gre) and bone morphogenetic protein (BMP); Fgf maintains Shh expression in the ZPA.

Negative feedback in the TGF- β family

The extensive TGF- β family of growth factors, which includes the activins, BMPs, GDFs, Nodal-related proteins and many others, controls a very large range of processes in development and growth control. A well conserved negative feedback loop regulates these important signals. The TGF- β s all signal through a class of serine/threonine kinase receptors and the signal is then transduced to the nucleus by the Smad family of proteins. These were first identified in *Drosophila*, and have since been found to be general TGF- β transducers. Receptor-binding Smads are phosphorylated and then complex with a co-Smad, Smad4, allowing them to transduce to the nucleus where they form transcriptional complexes (Fig. 8)^{64–66}. A subclass of inhibitory Smads has been identified in *Drosophila* (Dad) and vertebrates (Smad6 and Smad7)^{67–70}.

Although there is some controversy about how these inhibit signalling, in general the inhibitory Smads compete for binding either to the receptor (they lack a phosphorylation site) or to Smad4 (refs 64–66). Their expression is activated by TGF- β signalling, so they form a classic negative feedback loop. The conservation of this feedback loop from flies to humans emphasizes its significance, as does the discovery of tumours with elevated levels of the inhibitory Smads⁷¹ (TGF- β signalling generally inhibits proliferation).

Recently, another negative feedback loop relevant to TGF- β and cancer has been discovered. The SnoN oncoprotein has been shown⁷² to be a co-repressor that blocks Smad4-mediated transcriptional activation (Fig. 8). Upon TGF- β signalling, activated Smad3 enters the nucleus and relieves this repression, allowing target genes, which include *snoN* itself, to be transcribed. This SnoN negative feedback loop probably ensures that these target genes are kept stably off in the absence of significant TGF- β signalling. In its oncogenic form, SnoN is no longer susceptible to Smad3 de-repression, so it cannot escape from the negative feedback loop and TGF- β growth inhibition is lost.

Negative feedback generates stability

Some of the above examples illustrate one of the general consequences of negative feedback loops—the generation of systems that are stable even when confronted with environmental fluctuations⁷³. This property has recently been quantified in an artificially engineered gene network, where it was shown that negative feedback provided a twofold increase in stability over a broad range of input values⁷⁴. This homeostatic function of negative feedback has an important role in protecting cells from the uncontrolled growth and developmental aberrations that can lead to cancer following environmental damage.

p53 is a tumour suppressor whose activity is lost in the majority

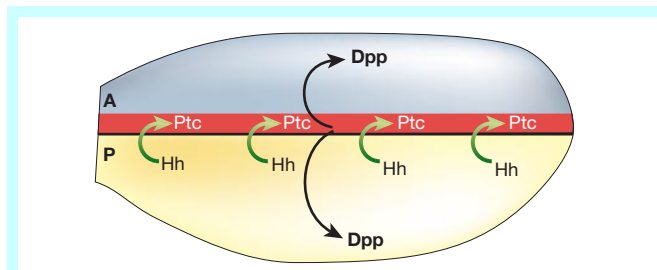


Figure 6 Variation on a negative feedback loop in the fly wing. The signalling protein Hedgehog (Hh) is expressed in all posterior (yellow) wing cells. The signal can only be transduced by anterior (grey and red) cells. High levels of Hh signalling induces the expression of the Hh binding protein, Patched (Ptc), in cells (red) near the Hh source. Patched sequesters Hh and prevents it from spreading further, causing Hh signalling to be restricted to cells close to the anterior/posterior (A/P) border. These cells express Dpp, which patterns the wing. The region in red is thus established as an organizing centre.

of human cancers. Its normal function is to induce cell-cycle arrest and/or cell death in damaged cells⁷⁵. In healthy cells, p53 is maintained safely below its functional threshold by the action of a negative feedback loop—it is inhibited by the protein Mdm2, whose own transcription is activated by p53 (refs 76, 77). When stressed, a cell needs to break out of this negative feedback loop, so that p53 can perform its cellular policing function. As with the inability to break out of the SnoN feedback loop discussed above, cells that cannot escape the Mdm2/p53 feedback loop are in danger of losing control of normal growth and development, as they cannot respond normally to potentially carcinogenic damage. Thus, a number of tumours contain amplified or overexpressed Mdm2. In these cases, the p53 gene is usually the wild type, indicating that the failure to escape from Mdm2 is sufficient to block the p53 response to cell damage, which leads to tumour progression.

Positive feedback loops can generate instability

In contrast to the stability produced by negative feedback, positive feedback can cause instability. In engineering applications, positive feedback is usually avoided as a dangerous phenomenon that can

lead to system failure (think of the famous Tacoma Narrows Bridge disaster). Nevertheless, organisms have evolved safeguards against this and I have discussed several essential and benign examples of positive feedback in developmental signalling. But a striking counter-example, which highlights the potential instability, is the auto-crine, positive feedback loops that occur in tumour formation and progression.

The Ras/MAP kinase pathway is activated by the EGF receptor, and it has been proposed that autocrine secretion of ligands is a significant cause of EGF receptor hyperactivity in cancer. Recently, EGF receptor ligands have been shown to be induced by Ras/MAP kinase signalling, providing a direct demonstration of this positive feedback cycle (A. Schulze and J. Downward, personal communication). Interestingly, this pathogenic strategy directly parallels the positive EGF receptor feedback loops discovered in normal development—an example of the widespread phenomenon of tumorigenesis ‘hijacking’ normal developmental programmes.

Canalization

At the beginning of the 20th century, the embryologist Hans Spemann observed that development was robust: small perturbations were corrected so as not to cause lasting damage. He referred to this correction mechanism as “double assurance”. In 1942, Waddington wrote about this developmental robustness at greater length and coined the term “canalization” to refer to what are really two separate but related phenomena⁸⁰. The first is the buffering potential of normal development, allowing animals to develop normally under a range of environmental conditions. The second

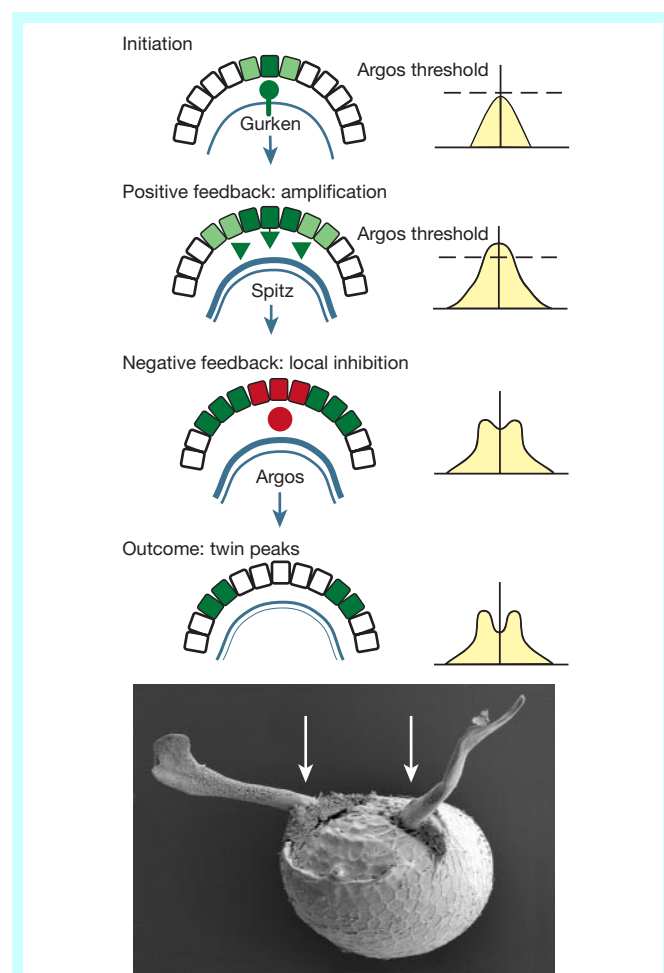


Figure 7 Integrated positive and negative feedback pattern in the *Drosophila* egg. EGF receptor signalling in the follicle cells is initiated by Gurken in the underlying oocyte. This initiates a positive feedback loop that amplifies EGF receptor signalling via the ligand Spitz (see Fig. 2). The amplified signalling then induces the expression of the inhibitor Argos, which blocks signalling near its source at the midline, thereby splitting the peak of signalling into two. The resultant twin peaks of EGF receptor signalling specify the lateral position of the prominent respiratory appendages, marked with arrows in the lower panel. The yellow curves indicate net EGF receptor signalling at each stage in the process.

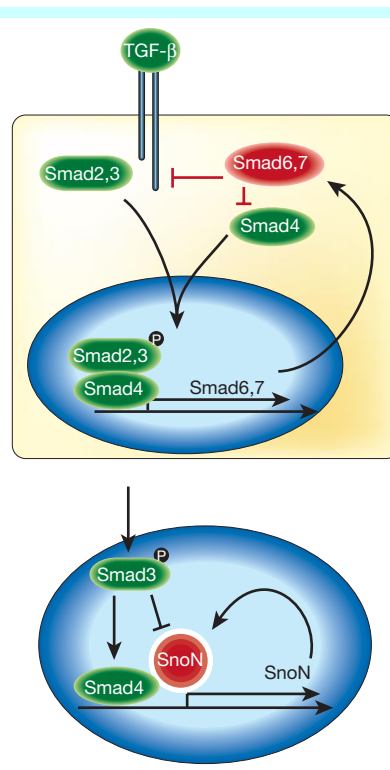


Figure 8 Two negative feedback loops that regulate TGF- β signalling and are implicated in cancer. Top panel, activated TGF- β receptors recruit and phosphorylate specific Smad proteins which, in complex with Smad4, activate transcription of target genes. These include the inhibitory Smad6 and Smad7 (Dad in *Drosophila*). TGF- β signalling inhibits cell proliferation, and, consequently, overexpression of the inhibitory Smads is associated with tumours. Bottom panel, the SnoN oncoprotein is a co-repressor that prevents transcriptional activation by Smad4 until phosphorylated Smad3 enters the nucleus. SnoN transcription is activated by Smad4, forming a negative feedback loop. Oncogenic SnoN is unable to break out of this feedback cycle as it is no longer sensitive to Smad3. P, phosphate.

is the robustness of developmental decisions—choices are amplified to all-or-nothing states rather than hovering around some intermediate.

These two phenomena lead to a marked invariance of the wild-type phenotype of animals. Redundancy of gene function after gene duplication during evolution provides one explanation for the relative stability of the wild-type state⁸¹. It is now, however, possible to suggest feedback mechanisms as another underlying principle that leads to canalization of development. And in fact the two aspects of the phenomenon—buffering and robustness—can neatly be explained by negative and positive feedback, respectively. Negative feedback allows self-adjustment of developmental signalling systems when they are perturbed⁸²; positive feedback causes initially small changes to become irreversibly large, the classic example being the amplification of small or even stochastic differences to select a single cell as a neural precursor in Notch-mediated lateral inhibition^{83,84}.

Perspectives

Over a rather short period of time, the importance of feedback control in normal development has become clear; it is also increasingly obvious that failure of feedback can lead to disease. Most pathways and a broad range of developmental decisions are regulated by feedback, both in vertebrates and invertebrates. It therefore seems reasonable to suggest that feedback loops provide a fundamental strategy for controlling cell signals in growth and development.

The recognition of negative feedback has paralleled the discovery in the past few years of a wide variety of signalling inhibitors. Although a reasonable proportion of these do participate in negative feedback control (that is, their expression is dependent on the signalling pathway they inhibit), there are also many that appear not to be regulated in this way. However, it can be difficult to determine whether a particular protein's activity is, directly or indirectly, controlled by a given signalling pathway, and it may turn out that more of these inhibitors than currently thought do participate in negative feedback. Negative feedback loops are relevant to disease as well as normal development. For example, the discovery of inhibitory Smad overexpression in tumours, and the p53/Mdm2 feedback loop are two examples of a theme that seems likely to become increasingly common.

Positive feedback, beyond its well-studied role in transcription factor autoregulation, has been less recognized but is nonetheless of real significance in development. In its simplest form, positive feedback can prolong and amplify the response to a weak signal. As described above, in its more intricate manifestations, it has a role in quite complex signalling outcomes, such as coordination of developmental events, defining precise domains of gene expression and pattern formation. As in engineering, however, positive feedback in biology always carries with it the risk of loss of control of signalling, and a subsequent breakdown of homeostasis that can lead to disease. □

- Mayr, O. The origins of feedback control. *Sci. Am.* **223**, 111–118 (1970).
- Weiss, P. & Kavanau, J. L. A model of growth and control in mathematical terms. *J. Gen. Physiol.* **41**, 1–47 (1957).
- Turing, A. M. The chemical basis of morphogenesis. *Philos. Trans. R. Soc. Lond. B* **237**, 37–72 (1952).
- Gierer, A. & Meinhardt, H. A theory of biological pattern formation. *Kybernetik* **12**, 30–39 (1972).
- Meinhardt, H. Biological pattern formation: new observations provide support for theoretical predictions. *Bioessays* **16**, 627–632 (1994).
- Smolen, P., Baxter, D. A. & Byrne, J. H. Mathematical modeling of gene networks. *Neuron* **26**, 567–580 (2000).
- Liu, K. D., Gaffen, S. L. & Goldsmith, M. A. JAK/STAT signaling by cytokine receptors. *Curr. Opin. Immunol.* **10**, 271–278 (1998).
- Naka, T. et al. Structure and function of a new STAT-induced STAT inhibitor. *Nature* **387**, 924–929 (1997).
- Starr, R. et al. A family of cytokine-inducible inhibitors of signalling. *Nature* **387**, 917–921 (1997).
- Endo, T. A. et al. A new protein containing an SH2 domain that inhibits JAK kinases. *Nature* **387**, 921–924 (1997).
- Nicola, N. A. et al. Negative regulation of cytokine signalling by the SOCS proteins. *Cold Spring Harbor Symp. Quant. Biol.* **64**, 397–404 (1999).

- Kovanen, P. E. & Leonard, W. J. Cytokine signalling: Inhibitors keep cytokines in check. *Curr. Biol.* **9**, R899–R902 (1999).
- Marine, J. C. et al. SOCS1 deficiency causes a lymphocyte-dependent perinatal lethality. *Cell* **98**, 609–616 (1999).
- Alexander, W. S. et al. SOCS1 is a critical inhibitor of interferon gamma signaling and prevents the potentially fatal neonatal actions of this cytokine. *Cell* **98**, 597–608 (1999).
- Marine, J. C. et al. SOCS3 is essential in the regulation of fetal liver erythropoiesis. *Cell* **98**, 617–627 (1999).
- Bjorbaek, C., Elmquist, J. K., Frantz, J. D., Shoelson, S. E. & Flier, J. S. Identification of SOCS-3 as a potential mediator of central leptin resistance. *Mol. Cell* **1**, 619–625 (1998).
- Adams, T. E. et al. Growth hormone preferentially induces the rapid, transient expression of SOCS-3, a novel inhibitor of cytokine receptor signaling. *J. Biol. Chem.* **273**, 1285–1287 (1998).
- Ray, R. P. & Schüpbach, T. Interleukin signaling and the polarization of body axes during *Drosophila* oogenesis. *Genes Dev.* **10**, 1711–1723 (1996).
- Wasserman, J. D. & Freeman, M. An autoregulatory cascade of EGF receptor signalling patterns the *Drosophila* egg. *Cell* **95**, 355–364 (1998).
- Morata, G. & Garcia-Bellido, A. *Wilhelm Roux Arch. Dev. Biol.* **179**, 125–143 (1976).
- Bienz, M. & Tremml, G. Domain of Ultrabithorax expression in *Drosophila* visceral mesoderm from autoregulation and exclusion. *Nature* **333**, 576–578 (1988).
- Thüringer, F. & Bienz, M. Indirect autoregulation of a homeotic *Drosophila* gene mediated by extracellular signalling. *Proc. Natl Acad. Sci. USA* **90**, 3899–3903 (1993).
- Bienz, M. Homeotic genes and positional signalling in the *Drosophila* viscera. *Trends Genet.* **10**, 22–26 (1994).
- Tremml, G. & Bienz, M. Induction of labial expression in the *Drosophila* endoderm: Response elements for dpp signalling and for autoregulation. *Development* **116**, 447–456 (1992).
- Gonzalez-Reyes, A., Macias, A. & Morata, G. Autocatalysis and phenotypic expression of *Drosophila* homeotic gene *Deformed*: its dependence on polarity and homeotic gene function. *Development* **116**, 1059–1068 (1992).
- McMahon, A. P., Joyner, A. L., Bradley, A. & McMahon, J. A. The midbrain-hindbrain phenotype of Wnt-1/Wnt-1- mice results from stepwise deletion of engrailed-expressing cells by 9.5 days postcoitum. *Cell* **69**, 581–595 (1992).
- Heemskerk, J., DiNardo, S., Kostrin, R. & O'Farrell, P. H. Multiple modes of engrailed regulation in the progression towards cell fate determination. *Nature* **352**, 404–410 (1991).
- Freeman, M., Klämbt, C., Goodman, C. S. & Rubin, G. M. The *argos* gene encodes a diffusible factor that regulates cell fate decisions in the *Drosophila* eye. *Cell* **69**, 963–975 (1992).
- Schweitzer, R., Howes, R., Smith, R., Shilo, B. -Z. & Freeman, M. Inhibition of *Drosophila* EGF receptor activation by the secreted protein Argos. *Nature* **376**, 699–702 (1995).
- Golembo, M., Schweitzer, R., Freeman, M. & Shilo, B. -Z. *argos* transcription is induced by the *Drosophila* EGF receptor pathway to form an inhibitory feedback loop. *Development* **122**, 223–230 (1996).
- Freeman, M. Reiterative use of the EGF receptor triggers differentiation of all cell types in the *Drosophila* eye. *Cell* **87**, 651–660 (1996).
- Freeman, M. Cell determination strategies in the *Drosophila* eye. *Development* **124**, 261–270 (1997).
- Perrimon, N. & McMahon, A. Negative feedback mechanisms and their roles during pattern formation. *Cell* **97**, 13–16 (1999).
- Ruohola-Baker, H. et al. Spatially localized rhomboid is required for establishment of the dorsal-ventral axis in *Drosophila* oogenesis. *Cell* **73**, 953–965 (1993).
- Golembo, M., Yarnitzky, T., Volk, T. & Shilo, B. Z. Vein expression is induced by the EGF receptor pathway to provide a positive feedback loop in patterning the *Drosophila* embryonic ventral ectoderm. *Genes Dev.* **13**, 158–162 (1999).
- Jin, M. H., Sawamoto, K., Ito, M. & Okano, H. The interaction between the *Drosophila* secreted protein argos and the epidermal growth factor receptor inhibits dimerization of the receptor and binding of secreted spitz to the receptor. *Mol. Cell. Biol.* **20**, 2098–2107 (2000).
- Casci, T., Vinós, J. & Freeman, M. Sprouty, an intracellular inhibitor of Ras signalling. *Cell* **96**, 655–665 (1999).
- Chiglione, C. et al. The transmembrane molecule kerkon 1 acts in a feedback loop to negatively regulate the activity of the *Drosophila* EGF receptor during oogenesis. *Cell* **96**, 847–856 (1999).
- Capdevila, J., Vogan, K. J., Tabin, C. J. & Izpisua Belmonte, J. C. Mechanisms of left-right determination in vertebrates. *Cell* **101**, 9–21 (2000).
- Schier, A. F. & Shen, M. M. Nodal signalling in vertebrate development. *Nature* **403**, 385–389 (2000).
- Nonaka, S. et al. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* **95**, 829–837 (1998).
- Sajih, Y. et al. Left-right asymmetric expression of lefty2 and nodal is induced by a signaling pathway that includes the transcription factor FAST2. *Mol. Cell.* **5**, 35–47 (2000).
- Gaio, U. et al. A role of the cryptic gene in the correct establishment of the left-right axis. *Curr. Biol.* **9**, 1339–1342 (1999).
- Yan, Y. T. et al. Conserved requirement for EGF-CFC genes in vertebrate left-right axis formation. *Genes Dev.* **13**, 2527–2537 (1999).
- Meno, C. et al. Mouse Lefty2 and zebrafish *antivin* are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell.* **4**, 287–298 (1999).
- Rodriguez Esteban, C. et al. The novel Cer-like protein Caronte mediates the establishment of embryonic left-right asymmetry. *Nature* **401**, 243–251 (1999).
- Yokouchi, Y., Vogan, K. J., Pearce, R. V. II & Tabin, C. J. Antagonistic signaling by *Caronte*, a novel *Cerberus*-related gene, establishes left-right asymmetric gene expression. *Cell* **98**, 573–583 (1999).
- Johnson, R. L. & Tabin, C. J. Molecular models for vertebrate limb development. *Cell* **90**, 979–990 (1997).
- Niswander, L., Jeffrey, S., Martin, G. R. & Tickle, C. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* **371**, 609–612 (1994).
- Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993–1003 (1994).
- Zuniga, A., Haramis, A. P., McMahon, A. P. & Zeller, R. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**, 598–602 (1999).
- Moon, A. M., Boulet, A. M. & Capocchi, M. R. Normal limb development in conditional mutants of Fgf4. *Development* **127**, 989–996 (2000).

53. Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by hedgehog protein. *Nature* **368**, 208–214 (1994).
54. Tabata, T. & Kornberg, T. B. Hedgehog is a signaling protein with a key role in patterning *Drosophila* imaginal discs. *Cell* **76**, 89–102 (1994).
55. Zecca, M., Basler, K. & Struhl, G. Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the *Drosophila* wing. *Development* **121**, 2265–2278 (1995).
56. Hooper, J. E. & Scott, M. P. The *Drosophila* patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765 (1989).
57. Ingham, P. W., Taylor, A. M. & Nakano, Y. Role of the *Drosophila* patched gene in positional signalling. *Nature* **353**, 184–187 (1991).
58. van den Heuvel, M. & Ingham, P. W. *smoothed* encodes a receptor-like serpentine protein required for hedgehog signalling. *Nature* **382**, 547–551 (1996).
59. Alcedo, J., Ayzenzon, M., Von Ohlen, T., Noll, M. & Hooper, J. E. The *Drosophila* *smoothed* gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* **86**, 221–232 (1996).
60. Chen, Y. & Struhl, G. Dual roles for *patched* in sequestering and transducing Hedgehog. *Cell* **87**, 553–563 (1996).
61. Marigo, V., Scott, M. P., Johnson, R. L., Goodrich, L. V. & Tabin, C. J. Conservation in hedgehog signaling: induction of a chicken *patched* homolog by Sonic hedgehog in the developing limb. *Development* **122**, 1225–1233 (1996).
62. Goodrich, L. V., Johnson, R. L., Milenkovic, L., McMahon, J. A. & Scott, M. P. Conservation of the *hedgehog/patched* signaling pathway from flies to mice: induction of a mouse *patched* gene by Hedgehog. *Genes Dev.* **10**, 301–312 (1996).
63. Chuang, P. T. & McMahon, A. P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **397**, 617–621 (1999).
64. Heldin, C. H., Miyazono, K. & ten Dijke, P. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).
65. Christian, J. L. & Nakayama, T. Can't get no SMADisfaction: Smad proteins as positive and negative regulators of TGF- β family signals. *BioEssays* **21**, 382–390 (1999).
66. Massague, J. TGF- β signal transduction. *Annu. Rev. Biochem.* **67**, 753–791 (1998).
67. Hayashi, H. *et al.* The MAD-related protein Smad7 associates with the TGF- β receptor and functions as an antagonist of TGF- β signaling. *Cell* **89**, 1165–1173 (1997).
68. Tsuneizumi, K. *et al.* Daughters against dpp modulates dpp organizing activity in *Drosophila* wing development. *Nature* **389**, 627–631 (1997).
69. Nakao, A. *et al.* Identification of Smad7, a TGF- β -inducible antagonist of TGF- β signalling. *Nature* **389**, 631–635 (1997).
70. Imamura, T. *et al.* Smad6 inhibits signalling by the TGF- β superfamily. *Nature* **389**, 622–626 (1997).
71. Kleeff, J. *et al.* The TGF- β signaling inhibitor Smad7 enhances tumorigenicity in pancreatic cancer. *Oncogene* **18**, 5363–5372 (1999).
72. Stroschein, S. L., Wang, W., Zhou, S., Zhou, Q. & Luo, K. Negative feedback regulation of TGF- β signaling by the SnoN oncoprotein. *Science* **286**, 771–774 (1999).
73. Bhalla, U. S. & Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387 (1999).
74. Becksei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
75. Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
76. Oren, M. Regulation of the p53 tumor suppressor protein. *J. Biol. Chem.* **274**, 36031–36034 (1999).
77. Ashcroft, M. & Vousden, K. H. Regulation of p53 stability. *Oncogene* **18**, 7637–7643 (1999).
78. Di Marco, E. *et al.* Autocrine interaction between TGF- α and the EGF-receptor: quantitative requirements for induction of the malignant phenotype. *Oncogene* **4**, 831–838 (1989).
79. Sporn, M. B. & Todaro, G. J. Autocrine secretion and malignant transformation of cells. *N. Engl. J. Med.* **303**, 878–880 (1980).
80. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563–565 (1942).
81. Wilkins, A. S. Canalization: a molecular genetic perspective. *BioEssays* **19**, 257–262 (1997).
82. Golembo, M., Raz, E. & Shilo, B. Z. The *Drosophila* embryonic midline is the site of Spitz processing, and induces activation of the EGF receptor in the ventral ectoderm. *Development* **122**, 3363–3370 (1996).
83. Heitzler, P. & Simpson, P. The choice of cell fate in the epidermis of *Drosophila*. *Cell* **64**, 1083–1092 (1991).
84. Collier, J. R., Monk, N. A., Maini, P. K. & Lewis, J. H. Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *J. Theor. Biol.* **183**, 429–446 (1996).

Acknowledgements

I am grateful to M. Bienz, T. Casci and S. Munro for their help with the manuscript.

Correspondence and requests for materials should be addressed to the author (e-mail: MF1@mrc-lmb.cam.ac.uk).

Ancient Egyptian chronology and the astronomical orientation of pyramids

Kate Spence

Faculty of Oriental Studies, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK

The ancient Egyptian pyramids at Giza have never been accurately dated, although we know that they were built approximately around the middle of the third millennium BC. The chronologies of this period have been reconstructed from surviving lists of kings and the lengths of their reigns, but the lists are rare, seldom complete and contain known inconsistencies and errors. As a result, the existing chronologies for that period (the Old Kingdom) can be considered accurate only to about ± 100 years, a figure that radiocarbon dating cannot at present improve. Here I use trends in the orientation of Old Kingdom pyramids to demonstrate that the Egyptians aligned them to north by using the simultaneous transit of two circumpolar stars. Modelling the precession of these stars yields a date for the start of construction of the Great Pyramid that is accurate to ± 5 yr, thereby providing an anchor for the Old Kingdom chronologies.

The pyramids of the middle of the third millennium in Egypt (the Old Kingdom), built as tombs for the kings of the period, were oriented to the cardinal points with extraordinary precision. The most accurately aligned is the Pyramid of Khufu (Cheops) at Giza, also known as the Great Pyramid, the sides of which deviate from true north by an average of less than 3 minutes of arc¹. Ever since modern survey techniques revealed this achievement in the late nineteenth century², the question of how the ancient Egyptians achieved such accuracy has been widely debated.

The absence of contemporary source material accounts for the range of possible orientation methods that have been proposed over the years. There are no relevant texts or representations from this period and discussions have therefore relied either on much later textual or representational evidence or on considerations of potential accuracy³. Although for many years it has been accepted on the grounds of accuracy that a stellar method was used^{3,4}, recent research has revived the possibility of solar alignment⁵.

Until now, discussions of ancient Egyptian orientation methods have overlooked the evidence provided by the alignments of the pyramids themselves. Researchers have made the implicit assumption that the ancient method used for orientation had the potential to produce the same degree of accuracy at any period³ but this assumption is not supported by the evidence. Lists of measurements of the alignment of pyramids (refs 3, 6) show that the alignment of

the pyramid of Khufu represented a peak of accuracy which was not maintained in subsequent reigns. In fact, after Khufu's reign the alignment of pyramids became increasingly inaccurate. If the Egyptians had mastered a method of exceptionally accurate orientation, they should have been able to reproduce the results in subsequent generations. Nor can the accuracy of the Khufu alignment be considered coincidental when overall trends in pyramid orientation are examined in detail.

Table 1 lists the pyramids for which accurate measurements of orientation are available together with the accession dates of the kings for whom they were constructed. As a result of differences in reconstructions of the historical data, several chronologies of the period are available⁷; here I follow the lower range of dates given in von Beckerath's recent chronology⁸, with the exception of the reign of Snofru. Stadelmann's proposed 46-year reign of Snofru is followed here⁹, and the start dates for Snofru's construction of Meidum and the Bent and Red pyramids also follow his chronology for the reign¹⁰.

I assume here that the pyramid alignment ceremony occurred in year 2 of each king's reign (with the exception of those for the later pyramids of Snofru), after the burial of his predecessor, the choice of a suitable location, and preparation and levelling of the site. This accords with the fact that even kings with short reigns are known to have started construction of their own tombs.

Table 1 Date and orientation of ancient Egyptian pyramids

Ruler	Currently accepted accession date	Orientation — west side of pyramid	Orientation — east side of pyramid	Recalibrated accession date
Djoser	2640 BC		~ + 180' (ref. 3)	
Snofru—Meidum (1)	2600 BC (−2/+17)	−18.1' (ref. 11) $\pm$ 1.0	−20.6' (ref. 11) $\pm$ 1.0	2526 BC $\pm$ 7
Snofru—Bent Pyramid (2)	[2583 BC] (−2/+11)	−11.8' (ref. 12) $\pm$ 0.2	−17.3' (ref. 12) $\pm$ 0.2	
Snofru—Red Pyramid (3)	[2572 BC] (−2/+9)		−8.7' $\pm$ 0.2	
Khufu (4)	2554 BC	−2.8' (ref. 1) $\pm$ 0.2	−3.4' (ref. 1) $\pm$ 0.2	2480 BC $\pm$ 5
Khafre (5)	2522 BC (−1)	−6.0' (ref. 1) $\pm$ 0.2	−6.0' (ref. 1) $\pm$ 0.2	2448 BC $\pm$ 5
Menkaure (6)	2489 BC (−4)	Average: +14.1' $\pm$ 1.8' (ref. 2)	+12.4' (ref. 2) $\pm$ 1.0	2415 BC $\pm$ 10
Sahure (7)	2446 BC (−15)		~ −23' (ref. 3, 6) $\pm$ 10	2372 BC $\pm$ 25
Neferirkare (8)	2433 BC (−16)		~ + 30' (ref. 3) $\pm$ 10	2359 BC $\pm$ 25
Unas	2317 BC	+ 17.4' (ref. 1)	+ 17.1' (ref. 1)	
Senwosret I	1956 BC		~ −90' (ref. 13)	
Amenemhat III	1853 BC		+ 15.7' (ref. 14)	

* Accurate measurement of alignment provided by J. Dörner.

Numbers in parentheses after rulers' names refer to labelling in Figs 1 and 4. All measurements of orientation are in arcminutes. Error margins are given only for those structures that appear in Figs 1b and 4. The dates in column 2 are from von Beckerath's chronology (lower estimates)⁸ with the exception of the length of Snofru's reign and the dates of construction of his pyramids (in square brackets) which follow Stadelmann^{9,10}. Error margins (in parentheses) given for these dates reflect cumulative differences in reign length between existing chronologies and are calculated relative to the beginning of Khufu's reign. For Snofru's reign a range of 29 to 48 years is allowed, reflecting current debate^{10,15}. For pyramids post-dating Khufu, the error margins reflect differences between two standard scholarly chronologies^{8,16}. Dates plotted in Figs 1b and 4 are two years later than those tabulated (with the exception of the Bent and Red pyramids) as it is assumed here that alignment ceremonies took place in year 2 of each king's reign. Error margins for the orientation of the west and east sides of the pyramids are estimates based on the equipment used to measure the alignments. ± 0.2 arcminutes is allowed for recent measurements taken with a meridian-seeking theodolite¹ and ± 1 for those measured using a less accurate theodolite and star-sightings^{2,11}. A nominal 10-arcminute allowance is made for the pyramids of Sahure and Neferirkare, the orientations of which were calculated from figures published in excavation reports³. The east side of Sahure's pyramid incorporates a surveying error, the size of which has been calculated from published measurements⁶. The final column shows dates recalibrated according to the method described in the text.

The dates and measurements of alignment for each pyramid are plotted in Fig. 1a. The figure shows that of the eight pyramids dating from 2600–2400 BC, six lie approximately in a straight line. The other two, the pyramids of Khafre and Sahure, lie close to this group. There is no clear relationship between this group of pyramids and those pre- and post-dating it: the pyramid of Djoser predates the first attempts to orient structures accurately to the cardinal points, whereas for the period from 2400–1800 BC we have only three structures for which measurements are available, too few to assess whether they form a pattern amongst themselves.

The eight pyramids constructed between the reigns of Snofru and Neferirkare are re-plotted in Fig. 1b. Researchers proposing stellar methods agree that the Egyptians used northern or circumpolar stars for orientation^{1,3,4}, which suggests that the alignment ceremony was carried out for either the east or west side of the pyramid (the central axis was probably never levelled, ruling out its orientation by this method). Where available, measurements of both east and west sides are marked. In some of the earlier pyramids there is considerable variation in the alignment of the different sides but this appears to be the result of difficulties in constructing right angles while laying out or extending the base. The graph suggests that only one side was accurately aligned and that it was the west side of the structure.

Figure 1b shows that if the pyramids of Khafre (number 5 in Fig. 1b) and Sahure (number 7) are temporarily ignored, the remaining six pyramids constructed between the reigns of Snofru and Neferirkare lie close to a straight line plotted through four of the points as a guide to the eye (line a). From this we can deduce that the pyramids were oriented by a method that varied in accuracy over time, becoming increasingly accurate until the reign of Khufu and then decreasing in accuracy at a steady rate. The fact that these six points converge closely on the line shows that the orientation technique could be applied with great precision despite the fact that the method itself was not accurate relative to true north.

Methods of orientation

These results are incompatible with any of the methods of orientation considered probable in current literature, solar or stellar, all of which involve the bisection of angles produced by measuring either

sun shadows or two equivalent positions on the trajectory of a northern or circumpolar star (isolated at equal height or in its most easterly and westerly positions)^{3–5}. These methods would have maintained a level of accuracy over time: graphically, all the measurements of alignment would lie in a band centred on true north. The width of the band would vary depending on the potential accuracy of the particular method and the points would be randomly scattered within this band. Experimental work has shown that a precise method of bisecting the angle between the most easterly and westerly positions of a northern star could potentially achieve an accuracy of within ± 3 arcminutes (J. Dorner, personal communication); however, only one measurement of actual pyramid alignment falls within a band of this width. Some of the more inaccurate methods proposed are likely to require error margins of more than ± 60 arcminutes.

In Fig. 1b, the apparent precision of the alignment method (shown by the proximity of the six points to line a) strongly suggests stellar orientation, as solar methods would be unlikely to produce such accurate results. This conclusion is supported by the progressive deviation of the alignments from true north, which could not be achieved using a solar orientation method. The most likely explanation of this deviation from north is that a stellar orientation method was used which became increasingly inaccurate as a result of the effects of precession.

The north celestial pole appears from the earth to be a point on the celestial sphere around which the stars rotate. This point is directly aligned with the axis of rotation of the earth. However, the revolving axis of the earth is itself unstable and rotates slowly, like a gyroscope. This is precession and, as a result, the north celestial pole appears to trace out a large circle on the northern sky, with each cycle taking around 26,000 years. This movement is extremely slow; without a system of recorded measurements it would be unnoticeable to observers, even over periods of hundreds of years.

Today, the position of the north celestial pole is marked approximately by the star α -Ursae Minoris (Polaris) which lies within one degree of the pole. However, during the period of pyramid construction there was no star accurately marking the pole. The closest star was α -Draconis but this lay nearly two degrees from the celestial pole at the beginning of the reign of Khufu. In any case, sightings

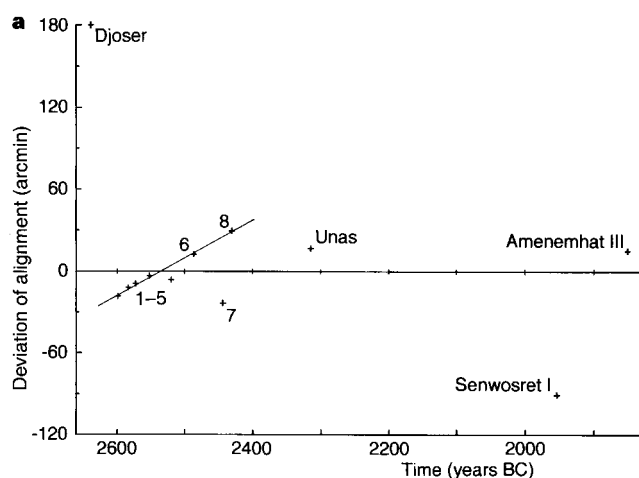
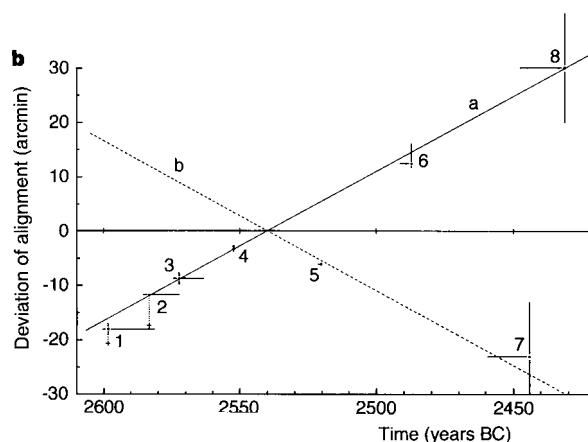


Figure 1 Deviation of pyramid alignments from true north over time. The dates plotted are 2 years later than the accession dates tabulated in Table 1 as it is assumed here that the alignment ceremony took place in year 2 of any given reign. 1, Meidum; 2, Bent Pyramid; 3, Red Pyramid; 4, Khufu; 5, Khafre; 6, Menkaure; 7, Sahure; 8, Neferirkare. **a**, 2640–1850 BC. Error bars are omitted as the uncertainties in some of the measurements of alignment are not well characterized; potential errors in the relative dating of the monuments are also difficult to assess over so large a period of time. **b**, 2600–2430 BC. West side alignments are plotted where available; east side alignments are marked as crosses below the west side measurement (numbers 1, 2, 4); additional error allowances



are made for the Red Pyramid (± 1) and Menkaure's pyramid (from calculations of the average and mean error of the other sides²) where no measurement of the west side alignment is available. Line a is plotted as a guide to the eye through the most recently and accurately surveyed points. If the simultaneous transit method was used to align the pyramids, the graph would form two lines of equal gradient, one positive and one negative, crossing the y-axis at 0. To show this, line b was derived from line a to be of equal but negative gradient; it passes close to the positions plotted for the pyramids of Khafre and Sahure. Pyramids lying on line a would be oriented with a star of Ursa Minor at its upper culmination and those on line b with a star of Ursa Major at its upper culmination.

taken toward a single star could not have achieved the results seen in Fig. 1b as, without a method of isolating culminations (the positions of greatest or least altitude of a star, which are also the points at which it crosses the meridian), the measurements would become increasingly random as the star moved away from the pole.

The method of orientation I propose is that the pole was considered to be located on an invisible chord linking two circumpolar stars on opposite sides of the pole. These two stars rotate around the pole, and when they are vertically aligned above the north horizon (one at its upper culmination and the other at its lower) an alignment made toward these stars with a plumb-line will be exactly oriented to true north, as long as the chord itself passes precisely through the pole. With the exception of the date when the chord between the stars lies exactly on the precessional trajectory of the pole, this method will produce alignments that become increasingly inaccurate at a steady rate over time. This method therefore has the potential to correspond with the results seen in Fig. 1b. This will be referred to as the simultaneous transit method because it uses two stars which cross (transit) the meridian simultaneously to establish true north.

Modelling the simultaneous transit method

A period from 2750 to 2350 BC (2550 BC $\pm$ 200 years, double the maximum error margin conventionally estimated for the relative chronologies of this period)⁷ was examined for pairs of bright stars within 15 degrees from the pole which could have been used in the simultaneous transit method. Using SkyMap Pro 6 (ref. 17) it was established that only two pairs of stars were joined by chords which crossed the celestial pole during this period: ζ -Ursae Majoris and β -Ursae Minoris (around 2467 BC), and ϵ -Ursae Majoris and γ -Ursae Minoris (around 2443 BC).

Figure 2a shows the circumpolar stars in 2467 BC, when a chord between ζ -Ursae Majoris and β -Ursae Minoris crossed the pole (the discrepancy between this date and the date of greatest precision of pyramid alignment shown in Fig. 1b will be discussed below). Figure 2b shows a view of the north horizon at Giza for the same date, showing the stars when they are vertically aligned (in simultaneous transit). A measurement of alignment taken with a plumb-line toward the stars at this time would be oriented exactly to true north.

The Khafre and Sahure alignments

The simultaneous transit method of alignment can also be used to explain the anomalous orientations of the pyramids of Khafre and Sahure, the two pyramids which do not conform with the trend set by the six other pyramids of the period. From Fig. 1b it is clear that although the deviation of these pyramids' alignment (numbers 5 and 7) lies west rather than east of north, they are approximately of the magnitude to be expected for their dates.

The simultaneous transit method involves making an alignment toward two stars when they are vertically aligned. The alignment can be taken when either of the stars is in the upper position. Because the upper culminations of the two stars fall approximately twelve hours apart, with the time of culmination of each star moving slowly through 24 hours in the course of a year, for about half the year the vertical alignment that falls within the hours of darkness will have one of the stars always above, and for the other half of the year the second star will be in the position of upper culmination when the two stars are vertically aligned during the course of the night.

When the chord between these two stars passes precisely through the celestial pole, an alignment taken toward the stars when they are

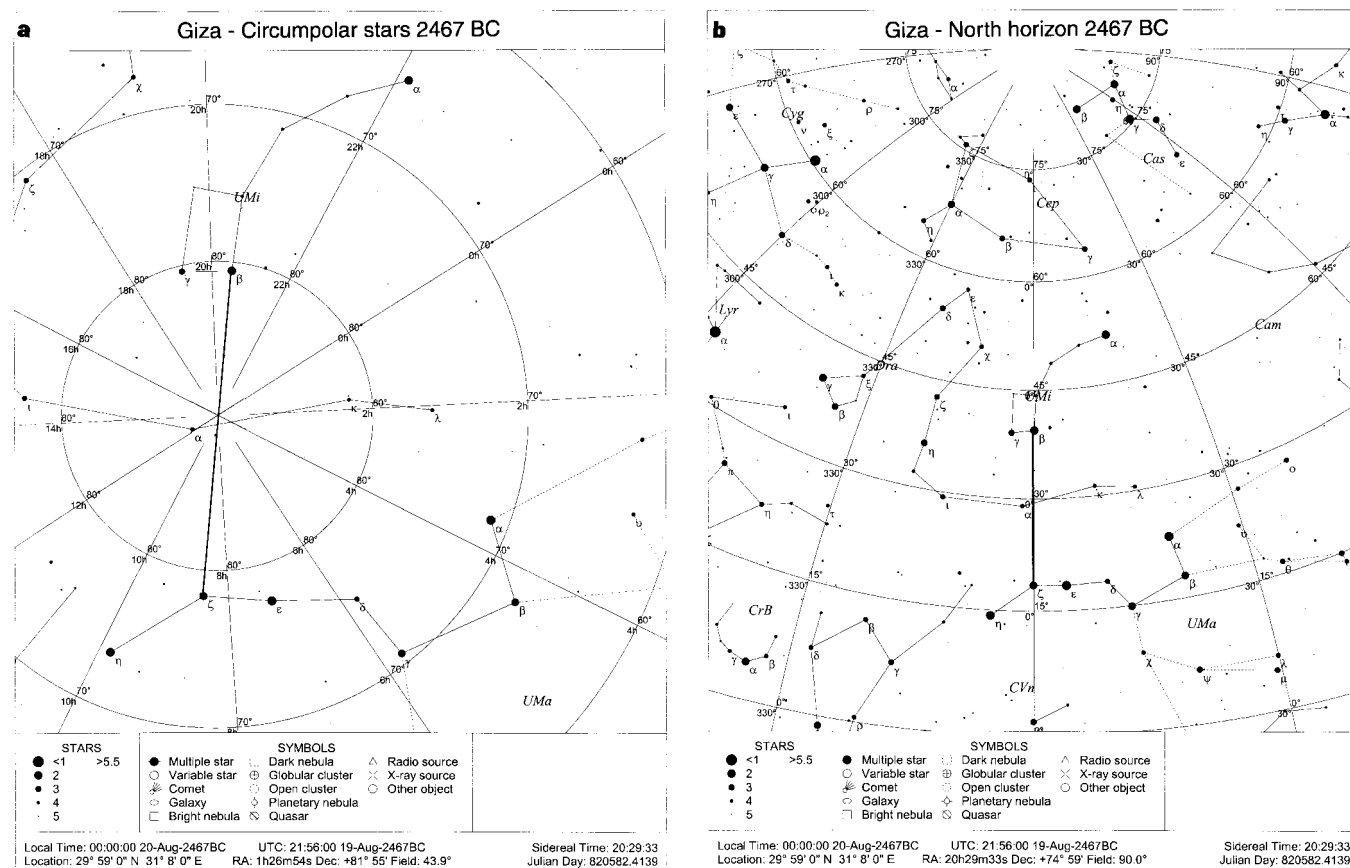


Figure 2 Modelling the simultaneous transit method for Giza, 2467 BC. **a**, A chord between stars β -UMi and ζ -UMa passes exactly through the north celestial pole. **b**, The

same stars in simultaneous transit. An alignment taken toward these stars using a plumb-line would be oriented exactly to true north. Maps produced on SkyMap Pro 6 (ref. 17).

in simultaneous transit will be exact with either star at its upper culmination. However, as the pole moves away from the chord between the stars as a result of precession, the chord itself begins to rotate around the pole. Alignments made using the simultaneous transit method will deviate from true north at a rate which increases steadily over time, regardless of which star is at its upper culmination when the alignment was made. However, as the chord itself is rotating around the pole, alignments made with a particular star at its upper culmination will be the opposite side of the north pole to alignments taken when the same star is at its lower culmination. This is illustrated in Fig. 3.

A graph of alignments made using the simultaneous transit method should therefore take the form of two lines of equal gradient, one positive and one negative, crossing at true north. In Fig. 1b, the gradient of line a is used to generate a second line, b, of equal but negative gradient, crossing line a at zero on the y-axis. This line passes close to the points plotted for the pyramids of Khafre and Sahure.

The alignment of Khafre's and Sahure's pyramids can therefore be shown to have been generated by the same method of alignment as the other six pyramids of the period, the only difference being the time of year at which the orientation ceremony was carried out, and therefore the star which was uppermost. To show this graphically, the pyramids of Khafre and Sahure can be re-plotted on the opposite side of true north (Fig. 4). Line a passes through all four points for which recent accurate measurements are available (numbers 2–5), through the error bars of the three later points and close to the point for number 1.

Modelling the results

In order to test further the theory that the simultaneous transit method was used to orient this group of pyramids, the method was modelled mathematically. F. R. Stephenson (personal communication) calculated the distance from the pole of a chord running between the stars ζ -Ursae Majoris and β -Ursae Minoris. He confirmed that the chord ran precisely through the pole in 2467 BC (astronomical date –2466) by computing the date when the right ascension of the two stars differed by exactly 180 degrees. The same procedure was repeated for the pair of stars ϵ -Ursae Majoris and γ -Ursae Minoris which produced the date 2443 BC (astronomical date –2442). He then computed the minimum angular distance between the north celestial pole and the great circle passing through the pairs of stars at 25-year intervals around these dates. The results are plotted in Fig. 4 (lines b and c) alongside the graph derived from archaeological data (line a).

Figure 4 shows that measurements of the angular distance of the pole from a chord running between stars ζ -Ursae Majoris and β -Ursae Minoris produce a line (b) of very similar gradient to that implied by the archaeological data. Stars ϵ -Ursae Majoris and γ -Ursae Minoris produce a significantly steeper gradient (line c) which, crucially, cannot be accommodated to the three Giza pyramids (numbers 4–6) which form the most accurately fixed group temporally and spatially. These results strongly support the hypothesis that the pyramids were aligned using the simultaneous transit method, which was carried out using stars ζ -Ursae Majoris and β -Ursae Minoris. The only major discrepancy between the results produced by the two independent data sets is time: the line

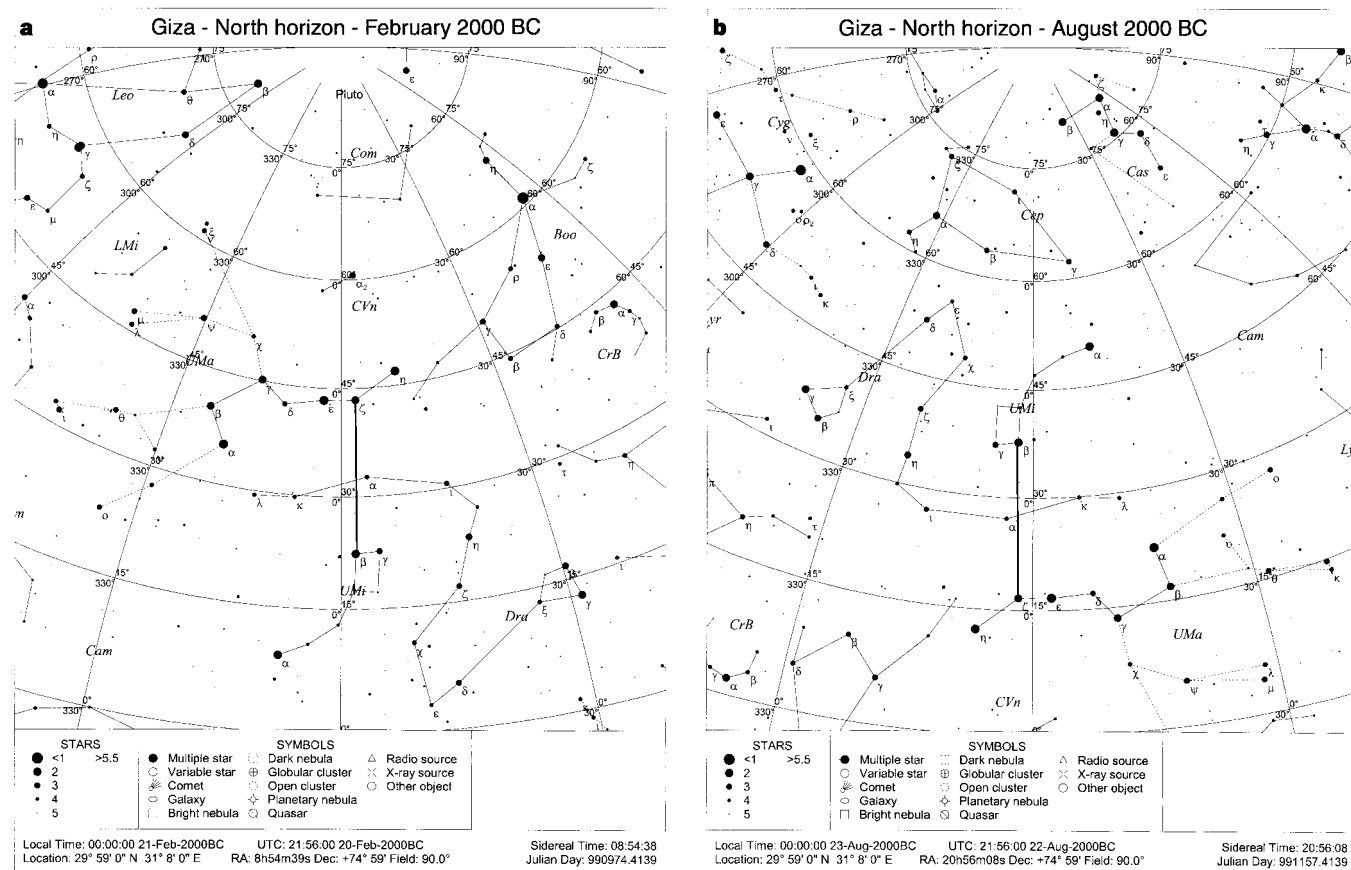


Figure 3 Deviation of alignments from true north resulting from use of the simultaneous transit method of orientation. As the chord between the pair of stars moves away from the pole, alignments taken toward the stars when they are in simultaneous transit (actually vertically aligned) may be on either side of the pole depending on which star is at its upper culmination. The size of the deviation from north will be the same at a given date

regardless of whether it is east or west of north. Here a much later date is mapped to exaggerate the effect of time on the result of the alignment process. **a**, Winter and spring alignments with ζ -Uma at upper culmination; resulting alignments are east of north. **b**, Summer and autumn alignments with β -Umi at upper culmination; resulting alignments are west of north. Maps produced on SkyMap Pro 6 (ref. 17).

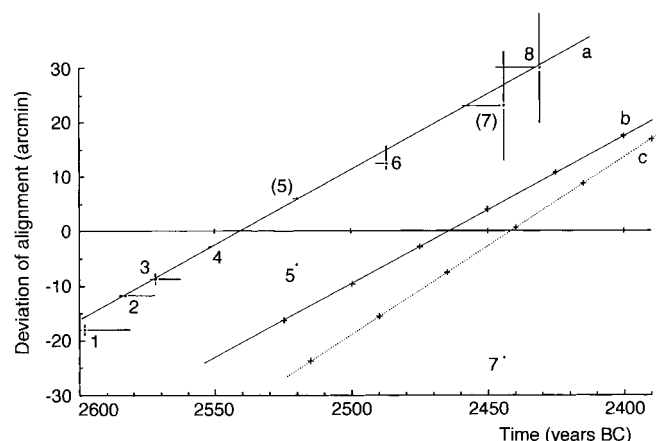


Figure 4 Astronomical modelling of the simultaneous transit method of orientation. Line a shows the deviation of pyramid alignment over time based on archaeological data plotted according to currently accepted chronologies of the period. Lines b and c plot the minimum angular distance between the north celestial pole and the great circle passing through pairs of stars which could have been used for simultaneous transit orientation (calculations by F. R. Stephenson). Line b uses stars β -UMi and ζ -UMa, which produce a gradient very similar to that implied by the archaeological data; line c uses stars γ -UMi and ϵ -UMa which produce a significantly steeper gradient. 1, Meidum; 2, Bent Pyramid; 3, Red Pyramid; 4, Khufu; 5, Khafre; 6, Menkaure; 7, Sahure; 8, Neferirkare. Parenthesized numbers (5 and 7) denote points re-plotted with positive rather than negative values, as described in the text, to conform to the dominant trend of the alignments.

generated by the astronomical modelling (line b) is over 70 years later than that derived from the archaeological results (line a).

Anchoring existing chronologies

The dating discrepancy between the two sets of results is caused by the fact that while stellar positioning at a given time can be predicted with great precision, existing Egyptian chronologies of this period based primarily on cumulative reign lengths can only be considered accurate to about ± 100 years⁷.

In Fig. 4, although the chronology of line b generated by astronomical data can be considered fixed, the chronology according to which the archaeological data are plotted (line a) is not anchored in time. However, the point at which line a crosses zero on the y-axis can now be fixed at 2467 BC from the results of the astronomical modelling. This gives a date of 2478 BC for the alignment of Khufu's pyramid which would require the lowering of von Beckerath's lower estimate of chronology by a further 74 years.

In reconstructing accession dates from dates of pyramid alignment ceremonies, potential for error theoretically exists in the assumption made here that this ceremony was held in the second year of each reign. However, it is exceptionally unlikely in this period that the error involved is more than ± 1 year. At present, a total of ± 5 years can be considered an adequate error allowance for

the reigns of Khufu and Khafre (F. R. Stephenson and T. van Albada, personal communications) given the accuracy of the archaeological data available for these reigns and the precision of the astronomical modelling.

Future research

The ability to fix the reigns of Khufu and Khafre to ± 5 years represents an advance in establishing a reliable absolute chronology for the second half of the third millennium BC in Egypt, but it does not solve all the problems. It is not possible simply to shift existing chronologies forward by the requisite number of years as fixed astronomical dates soon after 2000 BC mean that these existing chronologies will also have to be compressed. To achieve this, a process of careful reanalysis of the historical data will be necessary to make suitable adjustments. For this reason, the recalibrated accession dates given in the last column of Table 1 show error margins which increase over time as the possibility of numerous minor errors in cumulative reign lengths is compounded.

I intend to undertake fieldwork to collect more accurate data for those pyramids that have not been recently and reliably surveyed. From this and through more detailed mathematical modelling I hope to refine the error margin for dating the pyramids of Khufu and Khafre to ± 1 –2 years. More accurate data for the period around the reigns of Sahure and Neferirkare will reduce the error margins for the dates of these later kings and will assist in the process of refining the overall chronology of the period. □

Received 17 July; accepted 19 September 2000.

1. Dörner, J. *Die Absteckung und astronomische Orientierung ägyptischer Pyramiden* PhD thesis, Univ. Innsbruck (1981).
2. Petrie, W. M. F. *The Pyramids and Temples of Gizeh* (Field & Tuer, London, 1883).
3. Žába, Z. *L'orientation astronomique dans l'ancienne Égypte et la précession de l'axe du monde* (Editions de l'académie Tchecoslovaque des Sciences, Prague, 1953).
4. Edwards, I. E. S. *The Pyramids of Egypt* 3rd edn (Penguin, Harmondsworth, 1993).
5. Isler, M. An ancient method of finding and extending direction. *J. Am. Res. Center Egypt* **26**, 191–206 (1989).
6. Arnold, D. *Building in Egypt: Pharaonic Stone Masonry* (Oxford Univ. Press, Oxford, 1991).
7. Kitchen, K. The chronology of ancient Egypt. *World Archaeol.* **23**, 201–208 (1991).
8. von Beckerath, J. *Chronologie des pharaonischen Ägypten* (von Zabern, Mainz, 1997).
9. Stadelmann, R. *Die großen Pyramiden von Giza* (Akademische Druck- und Verlagsanstalt, Graz, 1990).
10. Stadelmann, R. Beiträge zur Geschichte des Alten Reiches. Die Länge der Regierung des Snofru. *Mitt. Deutsch. Archäologisch. Inst. Abteilung Kairo* **43**, 229–240 (1986).
11. Petrie, W. M. F. *Medum* (Nutt, London, 1892).
12. Dörner, J. Form und Ausmaße der Knickpyramide. Neue Beobachtungen und Messungen. *Mitt. Deutsch. Archäologisch. Inst. Abteilung Kairo* **42**, 43–58 (1986).
13. Arnold, D. *The Pyramid of Senwosret I* (Metropolitan Museum of Art, New York, 1988).
14. Arnold, D. *Der Pyramidenbezirk des Königs Amenemhet III. in Dahschur* (von Zabern, Mainz, 1987).
15. Krauss, R. The length of Sneferu's reign. *J. Egypt. Archaeol.* **82**, 43–50 (1996).
16. Baines, J. & Málek, J. *Atlas of Ancient Egypt* (Phaidon, Oxford, 1980).
17. Marriott, C. *SkyMap Pro Version 6* (Thompson Partnership, Uttroxteter, 1999).

<<http://www.skymap.com>>

Acknowledgements

For research funding I acknowledge the British Academy, the Wingate Foundation and the Lady Wallis Budge Fund, Christ's College, Cambridge. I thank all those who discussed this project with me, in particular B. Kemp; J. Dörner for allowing me to use measurements of orientation from his PhD thesis; F. R. Stephenson for astronomical calculations; T. van Albada and A. Egberts for their comments. I also thank N. Gay for technical assistance.

Correspondence should be addressed to K.S. (e-mail: kes1004@cam.ac.uk).

Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference

Andrew G. Fraser^{*†}, Ravi S. Kamath^{*†}, Peder Zipperlen^{*†}, Maruxa Martinez-Campos^{*}, Marc Sohrmann[‡] & Julie Ahringer^{*}

^{*} Wellcome/CRC Institute, University of Cambridge, Tennis Court Road, CB2 1QR Cambridge, UK

[‡] The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

[†] These authors contributed equally to this work

Complete genomic sequence is known for two multicellular eukaryotes, the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster*, and it will soon be known for humans. However, biological function has been assigned to only a small proportion of the predicted genes in any animal. Here we have used RNA-mediated interference (RNAi) to target nearly 90% of predicted genes on *C. elegans* chromosome I by feeding worms with bacteria that express double-stranded RNA. We have assigned function to 13.9% of the genes analysed, increasing the number of sequenced genes with known phenotypes on chromosome I from 70 to 378. Although most genes with sterile or embryonic lethal RNAi phenotypes are involved in basal cell metabolism, many genes giving post-embryonic phenotypes have conserved sequences but unknown function. In addition, conserved genes are significantly more likely to have an RNAi phenotype than are genes with no conservation. We have constructed a reusable library of bacterial clones that will permit unlimited RNAi screens in the future; this should help develop a more complete view of the relationships between the genome, gene function and the environment.

Complete genomic sequence is an invaluable tool for understanding the molecular mechanisms underlying an organism's development and function. The nematode worm *C. elegans* is one of two multicellular eukaryotes for which essentially complete genomic sequence is known^{1,2}. Thirty-six per cent of predicted *C. elegans* genes have a significant human match^{1,3} including many genes implicated in human disease^{3,4}, and functional analysis of the *C. elegans* genome has shed light on many conserved biological

processes and molecular pathways. A comprehensive functional analysis of all genes in *C. elegans* would greatly expand our knowledge of conserved gene function. We therefore decided to investigate systematically loss-of-function phenotypes of predicted genes of *C. elegans*, starting with chromosome I.

RNA-mediated interference (RNAi) transiently inhibits the activity of a gene by the introduction of double-stranded RNA (dsRNA) of sequence specific to the targeted gene⁵. The specificity and potency of RNAi make it ideal for investigating gene function beginning only with genomic sequence⁶. Ingestion of dsRNA-expressing bacteria results in RNAi of the targeted gene⁷, and we previously established that this technique is at least as effective as the injection of dsRNA for RNAi⁸: embryonic lethal phenotypes are detected with similar efficiency by feeding and injection, but feeding detects over 50% more post-embryonic phenotypes than injection. It is thus possible to make a library of dsRNA-expressing bacteria that can be used for high-throughput genome-wide RNAi screens at very low cost. It is important to note that because RNAi does not efficiently inhibit all genes, an RNAi-based screen will miss some

Table 1 Summary of phenotypes arising from RNAi of genes on chromosome I

Phenotype		Number	Per cent
All phenotypes	Total	339	13.9
Embryonic lethal	Emb	226	9.2
Sterile	Ste	82	3.4
	Stp	14	0.6
Developmental delay	Gro/Lva	145	5.9
Larval lethal	Lvl	38	1.6
	Unc	70	2.9
	Pvl	29	1.2
	Bmd	27	1.1
	Dpy	19	0.8
	Clr	14	0.6
Post-embryonic	Him	13	0.5
	Rup	9	0.4
	Mlt	8	0.3
	Pr2	8	0.3
	Egl	5	0.2
	Sck	5	0.2
	Bli	4	0.2
	Muv	2	0.1
	Rol	2	0.1
	Adl	1	<0.1
	Lon	1	<0.1
	Hya	1	<0.1

The number of predicted genes whose targeting through RNAi gave rise to each phenotype is shown. Percentages are given as percentage of total number of clones screened (2,445). Phenotypic classes were defined as described in Methods: the phenotypes are Emb (embryonic lethal), Ste (sterile), Stp (sterile progeny), Gro (slow post-embryonic growth), Lva (larval arrest), Lvl (larval lethality), Unc (uncoordinated), Pvl (protruding vulva), Bmd (body morphological defects), Dpy (dumpy), Clr (clear), Him (high incidence of males), Rup (ruptured), Mlt (molt defects), Pr2 (paralyzed), Sma (small), Egl (egg-laying defective) Sck (sick), Bli (blistering of cuticle), Muv (Multivulva), Rol (roller), Adl (adult lethal), Lon (long) and Hya (hyperactive).

Table 2 Detection of forward genetic loci on chromosome I by RNAi

Phenotype	Genetic loci fed	Possible to detect	RNAi phenotype detected	Published phenotype detected
All phenotypes	62	50	31	25
Embryonic lethal	21	21	19	16
Sterile	3	3	2	2
Sterile progeny	4	4	1	1
Developmental delay	0	0	—	—
Larval lethal	4	4	1	1
Post-embryonic	43	31	14	9

RNAi phenotypes were compared with those of genes that have known loss-of-function phenotypes. 'Genetic loci fed' denotes the number of genes in each category that were analysed by RNAi. 'Possible to detect' denotes the number of genes that have a loss-of-function phenotype that would have been detectable in our screen. 'RNAi phenotype detected' gives the number of genes for which a phenotype was identified. 'Published phenotype detected' gives the number of genes for which the RNAi phenotype matched a published phenotype. See Supplementary Information Table 2 for full data. Some of the differences between RNAi phenotypes and published phenotypes might be because RNAi can reduce both maternal and zygotic gene activities.

relevant genes. Despite this caveat, RNAi is a useful screening tool to complement classical forward genetics.

Analysis of the function of genes on chromosome I by RNAi

We constructed a library of bacteria expressing dsRNA corresponding to genes on chromosome I. Chromosome I is the second smallest chromosome, has few duplicated gene clusters and has no striking unusual features¹. Each individual bacterial clone is able to synthesize dsRNA designed to target a single gene; because gene predictions are still changing, a few primer pairs no longer

correspond to single genes (see Methods). In total, the resulting library contains 2,445 independent clones, corresponding to 2,416 predicted genes, a total of 87.3% of the 2,769 currently predicted genes of chromosome I.

We screened the library to identify genes whose inhibition gives a clearly detectable phenotype in wild-type worms (see Methods). We were able to assign a phenotype to 13.9% of the analysed genes, raising the number of sequenced genes on chromosome I with known phenotypes from 70 to 378 (Table 1). Many genes have more than one associated phenotype, reflecting that genes frequently have

Table 3 Partial list of RNAi phenotypes of genes on chromosome I

GenePairs	Locus	Description	Emb	Ste	P1	P2	P3	Dev	H
Chromosome dynamics									
B0511.8		CDC1-like	+	-	-	-	-	Lva	+
C37A2.4	cye-1	cyclin e	+	-	Clr	-	-	-	+
C41G7.2	*	kinesin	+	-	Him	-	-	-	+
C53H9.2		chrom stability	+	-	-	-	-	Gro	+
F57B10.12	mei-2	katanin homol	+	-	-	-	-	-	+
M01E11.6	*	kinesin	+	-	Him	-	-	Gro	+
R06C7.8		Bub1-like	+	-	Pvl	Rup	Lvl	-	+
T01G9.5	mei-1	ATPase	+	-	-	-	-	-	+
W09G3.3	*	RCC1 domains	-	-	Him	-	-	-	+
Y110A7A.d		cdc27 homologue	+	-	-	-	-	Gro	+
Y39G10A.246.e		MCM4-like	+	-	-	-	-	-	+
Y39G10A.246.i		INCENP-like	+	-	-	-	-	-	+
Y47G6A.247.i		pombe Rad2 homol	+	-	-	-	-	-	+
Y52B11A.9		Kin17 homol	+	-	Bmd	Rup	-	-	+
Cell structure									
C01G8.5		Ezrin-like	-	-	Unc	Lvl	-	Gro	+
C10H11.1		villin	-	-	Him	-	-	-	+
C17E4.9		PDZ domain	+	+	Unc	-	-	-	+
C32E8.10	unc-11	vesicle reg	+	-	-	-	-	-	+
C45G3.1		actin-binding	+	-	Unc	-	-	-	+
C47B2.3	* tba-2	tubulin	+	-	-	-	-	-	+
C53D5.a		nuclear import	+	+	-	-	-	Gro	+
C53D5.i		nuclear import	+	-	-	-	-	Gro	+
DY3.2	lam-1	nuclear lamin	+	+	Lvl	-	-	-	+
E03H4.8	*	beta coatomer-like	-	+	Unc	Stp	Clr	Gro	+
F07A5.7	unc-15	paramyosin	-	-	Unc	Prz	Egl	-	+
F20G4.3	nmy-2	non-muscle myosin	+	+	-	-	-	-	+
F21C3.5		MT nucleation	+	-	Unc	Bmd	-	Gro	+
F26B1.3		karyopherin	+	-	Lvl	-	-	-	+
F26E4.8	* tba-2	tubulin	+	-	-	-	-	-	+
F26H9.6		ras superfamily	+	+	Lvl	-	-	-	+
F28H1.2		calponin domain	+	-	-	-	-	Lva	+
F30A10.6		transporter	-	-	-	-	-	Gro	+
F36H2.1		cation transporter	+	-	-	-	-	Gro	+
F43G9.10		microfib assoc	+	-	-	-	-	Gro	+
F46F11.5		vacuolar ATPase	+	+	Lvl	-	-	-	+
F53B8.1	*	plectrins	+	-	Unc	Prz	Lvl	-	+
F53F10.5		nucleoporin-like	+	-	-	-	-	-	+
F54C1.7		tropoin c	-	+	-	-	-	-	+
F55A12.7		UNC-101 homol	-	-	Unc	-	-	Gro	+
F55F8.5		MT associated	-	-	-	-	-	Lva	+
F56F4.5		transporter	-	-	Him	-	-	-	+
H15N14.1		human NSF-like	+	+	-	-	-	-	+
M01A10.3		ribophorin	+	+	Lvl	Unc	-	-	+
R05D11.3		NTF2 homol	+	+	-	-	-	Gro	+
T03F1.9		UNC-89-like	+	-	-	-	-	-	+
T19B4.2		NUP153-like	+	+	-	-	-	-	+
T21E12.4	dhc-1	dynein heavy chain	+	-	-	-	-	-	+
T25G3.2		chitin synthases	+	-	-	-	-	Gro	+
T26E3.3	par-6	PDZ domain	+	-	-	-	-	-	+
W02B9.1	hmr-1	cadherin	+	-	Bmd	Unc	Dpy	-	+
W04C9.1		ABC transporter	+	-	-	-	-	Gro	+
Y105E8C.n		γ-adaptin	+	-	Unc	Lvl	-	Gro	+
Y18D10A.17		sup of clathrin defic	+	-	-	-	-	-	+
Y18D10A.20	pfn-1	profilin	+	-	-	-	-	-	+
Cell structure									
Y34D9A.152.a		vacuolar sorting	+	-	Unc	Prz	Lvl	Gro	+
Y48G8A.3945.e		adaptin subunit	-	-	Unc	-	-	-	+
Y71A12B.a		gravin-like	+	-	Dpy	Clr	-	-	+
Y71F9A.279.b		NXT1 homol	+	-	Pvl	Unc	-	-	+
Y71F9A.282.b		coatamer subunit	+	+	Unc	-	-	-	+
Y71F9A.290.a		NTF2 homol	-	-	Pvl	Rup	Clr	-	+
Y71G12A.195.e		talin	+	+	Unc	Prz	-	-	+
Y87G2A.s		HuVPS28 homol	-	-	Unc	-	-	-	+
Y87G2A.x		protein trafficking	-	-	Unc	-	-	Gro	+
Y87G2A.y		protein trafficking	-	-	Clr	-	-	-	+
ZK1014.1		vesicle fusion	-	+	Lvl	-	-	-	+
ZK1151.2		spectrin repeats	-	-	Unc	Bmd	Stp	-	+
Specific transcription									
B0025.3		txnl corepressor	+	-	-	-	-	Gro	+
C01G8.7		eyelid-like	+	+	Bmd	Lvl	-	Gro	+
C01G8.8		eyelid-like	+	-	Bmd	-	-	-	+
C01H6.5	nhr-23	nuc horm recep.	-	-	Unc	Lvl	Dpy	-	+
C12C8.3	lin-41	NHL domains	-	+	-	-	-	-	+
C32F10.7	nhr-2	nuc horm recep.	-	-	Him	-	-	-	+
C48E7.3		bZIP	-	-	-	-	-	Gro	+
D1081.2		SRF homol	-	-	Unc	Prz	-	-	+
F43G9.12		TCF-9-like	+	-	-	-	-	Gro	+
F52F12.6		MYT1 homol	-	-	Unc	-	-	-	+
F55F8.4		txnl repression	+	-	-	-	-	Lva	+
F57B10.1		bZIP	+	+	Sma	Dpy	-	Lva	+
K02B12.1	ceh-6	homeobox	-	-	Unc	Mit	-	-	+
M05B5.5	hlh-2	bHLH	+	+	Unc	Pvl	-	-	+
W02D3.9	unc-37	groucho family	+	+	Unc	-	-	-	+
W03D8.4	pop-1	HMG box	+	+	-	-	-	-	+
Y40B1A.4		Zn finger	-	-	Unc	Bmd	-	-	+
Y54E5B.3		Mediator complex	+	-	Rup	-	-	-	+
Y65B4A.179.b		txnl activator	-	-	Unc	Dpy	-	-	+
ZC247.3	lin-11	LIM homeodomain	-	-	-	-	-	Gro	+
ZK858.4	mel-26	kruppel-like	+	-	-	-	-	-	+
Signalling									
C09D1.1	unc-89	multiple domains	-	-	Unc	-	-	-	+
C10H11.9	let-502	ROCK	+	-	Dpy	Rol	Lvl	-	+
C26C6.2	goa-1	G-a subunit	+	+	Unc	Pvl	Egl	-	+
C32E8.5		FHA domain	+	+	-	-	-	Gro	+
F26E4.1		PP2A reg subunit	+	-	Lvl	-	-	-	+
F55A12.3		PI-4P 5' kinase	-	+	-	-	-	-	+
F55C7.4	unc-73	GEF	+	-	-	-	-	-	+
F55C7.7	unc-73	GEF	-	-	Egl	-	-	-	+
K04G2.8	apr-1	APC homol	-	-	Unc	Bmd	Lvl	-	+
K05C4.6	hmp-2	β-catenin	+	-	Unc	Dpy	Bmd	-	+
K12C11.2		SUMO-1 like	+	-	Pvl	-	-	-	+
T01G9.6	kin-10	CKIIβ subunit	+	-	Pvl	-	-	Gro	+
T21E3.1	*	PTPase	+	+	-	-	-	-	+
T23D8.1	mom-5	frizzled-like	-	-	Unc	Bmd	-	-	+
T23H2.5	*	rab family	-	-	Sma	Clr	-	-	+
Y106G6E.6		Casein Kinase 1	+	-	Dpy	-	-	Gro	+
Y18D10A.5	gsk-3	GSK-3	+	-	-	-	-	-	+
ZC581.1		NIMA-like kinase	-	-	Unc	Lvl	Rol	-	+
ZK265.6		G prot coup recep	-	-	-	-	-	Gro	+

RNAi phenotypes are shown for genes in the following functional classes: chromosome dynamics and cell cycle; cell structure; specific transcription; and signal transduction. For each gene, the following data are shown: the Research Genetics GenePairs name; whether a number of paralogs might be targeted (asterisk in column 2; see methods for criterion); the corresponding genetic locus name if it exists; a short description of gene function; the RNAi phenotype in which embryonic lethality (Emb), fecundity (Ste), post-embryonic phenotypes (P1–3) and developmental delay (Dev) are shown separately. Emb and Ste are classified into weak (white box, black '+') or strong (black box, white '+') phenotypes. For Emb, weak is 10–80% embryonic lethality, strong is 90% embryonic lethal or more; weak Ste denotes a broad size of 1–10, whereas strong Ste is totally sterile. Column H shows whether there is a match (white box, black '+') or a homologue (black box, white '+') in *Drosophila melanogaster*, *Saccharomyces cerevisiae* or humans. Phenotypic abbreviations are given in legend to Table 1. The GenePairs name does not always correspond with the current predicted gene name as gene predictions change.

several functions in the organism. Furthermore, as we examined both worms that were exposed to dsRNA only as larvae or adults as well as their progeny, we could assign post-embryonic phenotypes to genes that result in sterility or produce 100% embryonic lethal progeny. A summary of these results and a partial listing of the phenotypes obtained are given in Tables 1 and 3. Full results are given in Supplementary Information Table 1 and are publicly accessible in WormBase (<http://www.wormbase.org>).

Our screen was sufficiently effective to identify 90% of known embryonic lethal genes. In addition, we were able to assign phenotypes to 45% of genes with a known post-embryonic phenotype that should have been detectable in our screen (Table 2; and Supplementary Information Table 2). However, we failed to find phenotypes for a number of previously characterized genes. In some cases (for example, *fog-3*), this was not due to an inherent difficulty in inhibiting the genes using RNAi (as we obtained the correct phenotype in a separate experiment), but simply because we overlooked them in the screen. However, only one of eight genes involved in neuronal function gave a detectable RNAi phenotype; this agrees with our finding that neurons appear to be more resistant to RNAi than are other cell types⁸. Similarly, we did not detect phenotypes for several genes involved in sperm development (*fer-1*, *spe-9* and *spe-11*).

The largest phenotypic class, comprising over 60% of the genes, are those whose inhibition by RNAi gives rise to embryonic lethality—the Emb genes; these include a large number of components of the basal cellular machinery. More notably, we find a homologue of the SMN human disease gene⁹, a variety of genes encoding RNA-binding proteins (several such proteins have a role in early polarity; reviewed in ref. 10), a number of genes involved in chromosome condensation and separation, components of signal transduction pathways and many conserved genes that have no known biochemical function.

The largest class of post-embryonic phenotype is the Uncoordinated (Unc) class. Unc phenotypes arise from defects in the development or function of the neuromuscular system (reviewed in ref. 11). We find Unc genes encoding proteins involved in vesicle sorting and fusion as well as transcription factors (including a homologue of the zinc finger transcription factor MYT-1 which is only expressed in developing neurons in mammals^{12–14}) and components of the cytoskeleton (for example a kakapo^{15–18} and a talin¹⁹ homologue).

A number of genes showed a high incidence of males (Him)

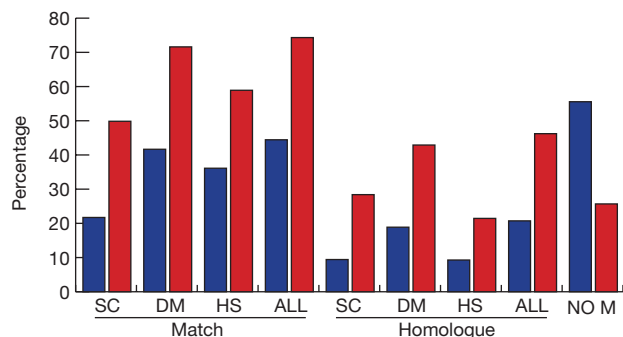


Figure 1 Conservation of genes with an RNAi phenotype. Matches or homologues of *C. elegans* genes were identified as described in the text. Shown are percentages of all genes (blue bars) or genes with RNAi phenotypes (red bars) with matches or homologues in *S. cerevisiae* (SC), *D. melanogaster* (DM), human (HS), or all three combined (ALL), or with no matches in any organism (NO M). The significance of the differences between the percentages of genes and the percentages of genes with RNAi phenotypes that have homologues is $P < 0.001$ in all cases, except for comparison with human homologues for which it is $P < 0.1$.

phenotype. *C. elegans* is usually grown as a self-fertilizing hermaphrodite with males arising at a low frequency in wild-type cultures owing to non-disjunction of the X chromosome (hermaphrodites have two X chromosomes, males only one). An increased number of males is indicative of either the incorrect segregation and maintenance of chromosomes in the germ line (reviewed in ref. 20) or defects in sexual specification. The Him genes that we identified include kinesins, a katanin homologue^{21,22} and a nuclear hormone receptor.

Conservation of genes with RNAi phenotypes

We examined the level of cross-species conservation of the genes for which we detected an RNAi phenotype (Fig. 1). To find *C. elegans* genes that are conserved in other species, we identified *C. elegans* genes that have hits with BlastP²³ expectation values below 1.00×10^{-6} in *Saccharomyces cerevisiae*, *Drosophila melanogaster* or humans; we defined these as a 'match'. Hits with BlastP e-values

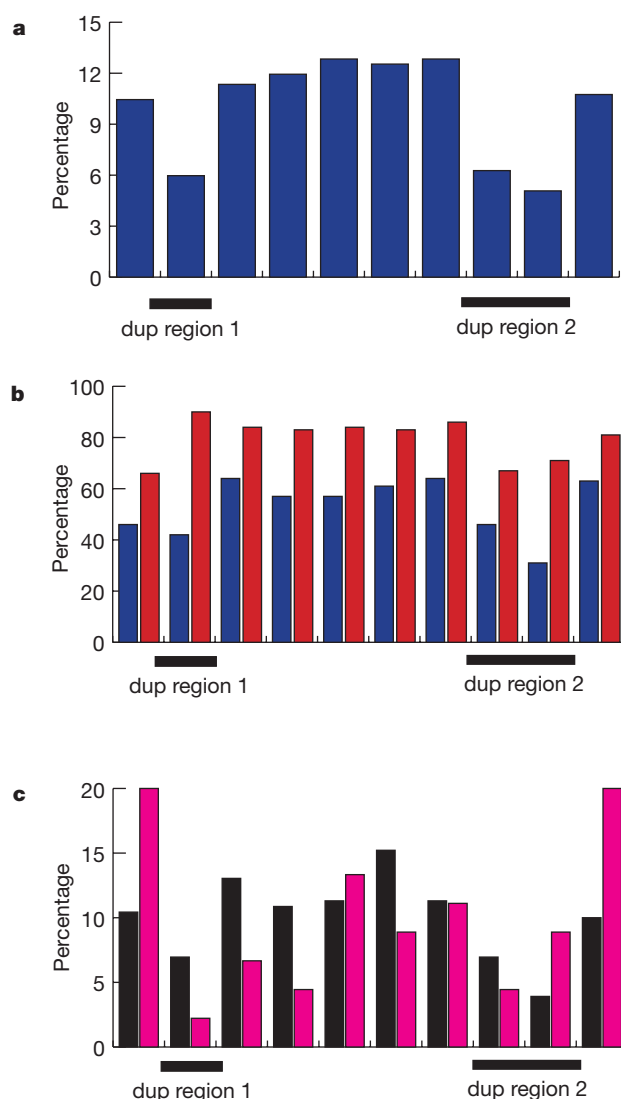


Figure 2 Distribution on chromosome I of genes with RNAi phenotypes and genes with ESTs. In each panel, chromosome I was analysed in 10 consecutive portions, each containing 10% of predicted genes. **a**, Percentage of all genes with an RNAi phenotype that are in each portion. **b**, Percentage of all predicted genes that have an EST (blue bars) or of genes that gave an RNAi phenotype that have an EST (red bars) in each portion. **c**, Percentage of Emb genes (black bars) or genes with viable post-embryonic phenotypes (pink bars) in each chromosomal portion. The 'dup region' indicates the approximate location of regions containing local duplications.

below 1.00×10^{-10} and in which the conservation extends over at least 80% of the *C. elegans* protein length, we defined as 'homologues'; this category includes orthologues. This provides a conservative estimate of the number of genes with regions of conservation (matches) or homologues, respectively.

We found that genes with RNAi phenotypes were much more likely to have a match ($P < 0.001$) compared with all genes (Fig. 1). Most striking is the similarity that we see between *C. elegans* and *Drosophila*: whereas 42% of *C. elegans* genes have a match and 19% have a homologue in *Drosophila*, over 72% of genes with an RNAi phenotype have a *Drosophila* match and 43% have a homologue (Fig. 1). This analysis shows that genes with a required function in *C. elegans* have been highly conserved across eukaryotic evolution. We also find that highly conserved genes are more likely to have an RNAi phenotype than genes that show no conservation: 26% of *C. elegans* genes that have a homologue in one of the organisms examined give an RNAi phenotype, compared with only 5% of genes with no conservation ($P < 0.001$).

Distribution of genes with RNAi phenotypes

Genes for which we identified an RNAi phenotype are evenly distributed across the chromosome with the exception of two regions (corresponding to segments 2 and 8–9 in Fig. 2a) for which there appears to be a drop in number. These two regions correspond to the two regions of chromosome I that contain locally duplicated gene clusters¹. We suggest that the reduction in the number of phenotypes observed by RNAi in these regions may be due to gene duplication and thus redundancy of function. It is worth noting that some of the predicted genes in the duplicated regions may not be expressed: whereas genes with RNAi phenotypes are equally likely to have an expressed sequence tag (EST) in all regions of the genome (see below), there is a significant drop ($P < 0.05$) in the proportion of total genes with ESTs in the second locally duplicated gene cluster region (Fig. 2b; 39% of genes in the

second cluster have an EST compared with 53% over the entire chromosome). We suggest that a portion of the predicted genes in such regions of duplication may in fact be pseudogenes.

Genes that give RNAi phenotypes are much more likely to have an EST than genes on chromosome I in general (82% compared with 53% respectively, $P < 0.001$; Fig. 2b). The relatively high percentage of genes with RNAi phenotypes that have ESTs may reflect that these genes are expressed at higher levels. It may also be that many genes that lack ESTs are only expressed conditionally; we are unlikely to have found phenotypes for such genes.

In *C. elegans*, there is evidence of differences between the chromosome arms and the central regions (the clusters), suggesting that there might be differences in gene type or function across the chromosome²⁴. In general, the distribution of genes in any given phenotypic class was similar to that for all genes with an RNAi phenotype (such as Emb genes; compare Fig. 2c and a). However, genes with viable post-embryonic phenotypes (Pep genes)—those that gave a post-embryonic phenotype without any embryonic or post-embryonic lethality, sterility or developmental delay—show a trend toward enrichment at the arms of chromosome I. It has been suggested that the chromosome arms may be more prone to mutation and recombination than the central core portion²⁴ and, if so, that novel gene functions are more likely to evolve in such regions. Our finding that genes which uniquely affect post-embryonic development cluster at the arms supports this model.

Biochemical function and RNAi phenotype

To explore the relationship between the biochemical function of a gene product and its mutant phenotype, we categorized the sterile (Ste), embryonic lethal (Emb), uncoordinated (Unc) and viable post-embryonic phenotype (Pep) genes into the functional classes shown in Fig. 3a.

Unsurprisingly, genes involved in basal metabolic processes account for ~50% of Ste and Emb genes (Fig. 3a); this confirms

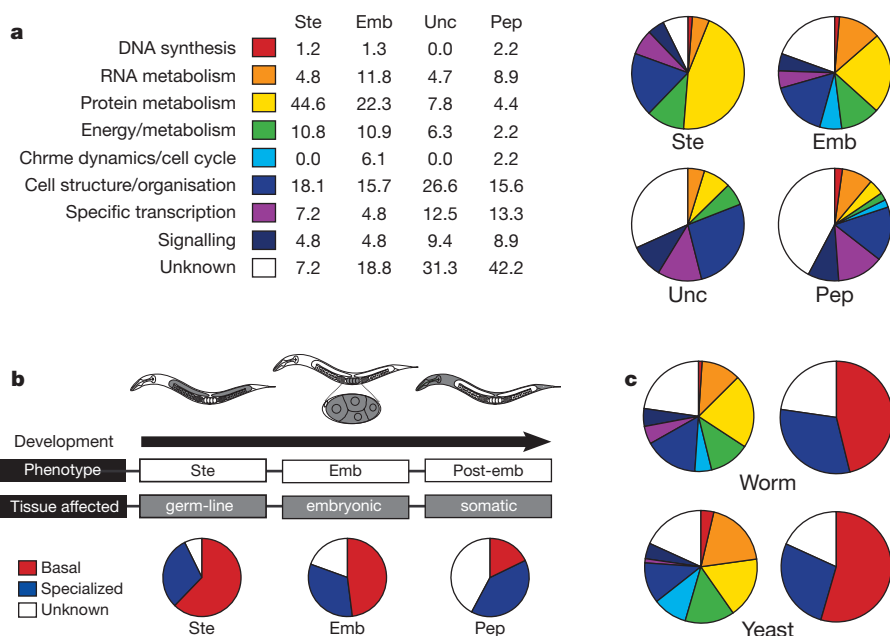


Figure 3 Functional classes of Emb, Ste, Unc and Pep genes. **a**, Predicted products of genes that gave Ste, Emb, Unc or viable post-embryonic (Pep) RNAi phenotypes were placed into functional classes as described in Methods. Genes whose products could not be accurately classified into any of the eight functional classes were placed into the unknown category (white). Numbers denote the percentage of genes in each functional class; pie charts illustrate these numbers graphically. **b**, Pie charts show distributions of predicted gene products grouped as follows: basal metabolic category (red) comprises the

classes of DNA, RNA, protein and intermediate metabolism; specialized functions (blue) comprises cell-cycle and chromosome dynamics, cell biology and cellular structure, gene-specific transcription factors and signal transduction. Worms show the tissue affected in each phenotypic class shaded in grey. **c**, Distribution of genes giving rise to non-viable RNAi phenotypes in *C. elegans* (worm) or to non-viable phenotypes following disruption in *S. cerevisiae* (yeast).

that these basic biochemical processes are indeed essential for viability. In contrast, under 20% of Unc and Pep genes encode components of the basal metabolic machinery, whereas more than twice as many encode proteins with more specialized functions (Fig. 3a, b). There is thus a clear difference between the types of gene required for germline function or embryonic viability (which mainly require basal machinery) and those involved in later developmental processes which appear to require proteins of either more specialized functions or as yet unknown function (Fig. 3b).

A second clear trend is that the number of genes of unknown function increases greatly in the Unc and Pep genes, making this the largest overall class for those phenotypes (Fig. 3). This shift underlies the fact that, although we know a great deal about basic metabolic processes of eukaryotic cells (and thus can readily ascribe function to a large proportion of Ste and Emb genes), much is still to be learnt about the complex processes and the genes that regulate the development and function of a multicellular eukaryote. A significant number (~25%) of genes of unknown function have close homologues in *Drosophila* or humans; further study of these may shed light on conserved processes specific to animals.

Essential genes of *S. cerevisiae* and *C. elegans*

Saccharomyces cerevisiae was the first eukaryote to be completely sequenced²⁵ and reverse genetics has been used extensively to investigate *S. cerevisiae* gene function. In a set of 3,680 genes knocked out by targeted disruption, 890 affect viability²⁶; we compared these genes to those that gave different RNAi phenotypes in *C. elegans*. Yeast and worm genes important for viability have a similar distribution within the different functional classes, but are different from the Unc or Pep distributions (Fig. 3c; also compare 3a and b). This suggests that similar types of gene are required for viability of yeast and animal cells. A striking difference ($P < 0.001$) is that only ~1% of the genes required for viability in yeast are transcription factors, whereas for *C. elegans* it is ~4% (a similar percentage of the genomes of yeast²⁷ and *C. elegans*² encode transcription factors, 3.3% and 2.5%, respectively). This suggests that many of the *C. elegans* transcription factors required for viability may be involved in developmental processes.

Estimate of the size of functionally non-redundant genome

What do our data tell us about the size of the functionally non-redundant genome? We screened 12.7% of the *C. elegans* genome and found that 339 genes gave a clearly discernible phenotype. Taking into account the sensitivity of our screen and scaling up to the entire genome, we estimate that ~5,400 genes will be individually required for wild-type *C. elegans* development under standard laboratory conditions (~2,300 genes for embryonic viability and ~3,100 post-embryonically; see Methods for calculation). This is comparable to previous estimates based on forward genetics²⁸. We expect that phenotypes for other genes will be identified under different conditions (for example, environmental stress), in other genetic backgrounds, or using more refined and restricted screening conditions.

Discussion

We have taken a systematic approach to identify functions for the predicted genes of *C. elegans* chromosome I. This is the first large-scale reverse genetic analysis of a multicellular organism and has increased by fivefold the number of sequenced genes with known phenotypes on this chromosome.

Although we have identified RNAi phenotypes for many genes, some will have eluded our screen for one of at least two reasons. First, RNAi may have been ineffective against the targeted gene. RNAi does not accurately phenocopy the null phenotype of all genes (such as genes involved in neuronal function), and may result in either partial or no loss of function. It should also be noted that if several genes have regions of identical or near-identical nucleotide

sequence, RNAi could target them simultaneously, so that the observed phenotype may be the result of the inhibition of more than one gene. Second, we will not have detected either subtle or conditional phenotypes. However, we anticipate that future RNAi-based screens using specific assays should be able to detect phenotypes for many more genes, thus increasing our understanding of *C. elegans* and hence of metazoan biology in general. As our library consists of bacterial clones that can be replicated, and the feeding protocol is relatively simple compared with injection, the library can be used repeatedly at low cost and high efficiency for such screens. In addition, we expect that a feeding library and database of associated phenotypes will prove valuable for the positional cloning of genes; currently there are over 300 genes on chromosome I identified by mutation but not yet cloned.

Although the time needed for an RNAi screen using our bacterial library is similar to that for a classical genetic screen, the two approaches have different advantages and will yield different results. Both approaches can be used to screen the entire genome for genes involved in a particular process, and both may identify complete or partial loss-of-function phenotypes. Classical forward genetics generates stable mutant lines that can be maintained indefinitely; furthermore, whereas some genes are resistant to RNAi, all genes are sensitive to mutagens (albeit to a greater or lesser degree) and could thus be cloned using a classical screen. Also, some mutants isolated by forward genetics are due to gain-of-function mutations, which cannot be generated by RNAi. However, the positional cloning of a gene is often slow and laborious. RNAi has the disadvantages mentioned above, but it also has the key advantage of all reverse genetics: the sequence of the gene is already known, and thus any mutant phenotype observed is automatically connected to a known sequence.

In the future, we aim to extend our library construction and functional analysis to the entire *C. elegans* genome and anticipate that the possibility of genome-wide RNAi screening, in conjunction with other functional genomics approaches such as expression analyses using microarrays²⁹ and two-hybrid experiments³⁰ will accelerate *C. elegans* research. □

Methods

Generation and cloning of PCR products

PCR products were synthesized using BioTaq polymerase (Bioline) in a reaction containing 25 ng of *C. elegans* genomic DNA, 20 pmol of *C. elegans* GenePairs primers (Research Genetics) and 100 µM dNTPs: 34 cycles of (94 °C 30 s, 58 °C 30 s, 72 °C 90 s) were followed by an extension of 1 h at 72 °C to enhance A-tailing of products. Products were ligated into linearized T-tailed L4440 vector⁷ and transformed into the HT115(DE3) bacterial strain (L. Timmons and A. Fire, personal communication) using standard methods. Colonies containing correct sized insert were identified by PCR using vector-specific oligos, and the cloned inserts confirmed by PCR using the original Research Genetics primer pair. Primer sequences are available at <http://cmgm.stanford.edu/~kimlab/primers.12-22-99.html>.

RNAi screening

RNAi was performed essentially as described in ref. 8, where feeding data were reported on 86 of the 2,445 genes described here. Briefly, 4 wells of a 12-well plate containing NGM agar + 1 mM IPTG + 25 µg ml⁻¹ carbenicillin were inoculated with bacterial cultures grown 8–18 h for each targeted gene. Between 10 and 15 L3–L4 stage worms were placed in the first of the 4 wells for each gene and left for 72 h at 15 °C. Three worms, now young adults, were removed and individually placed on three remaining wells for each gene and allowed to lay embryos for 24 h at room temperature; the three worms were then removed ($t = 0$). The phenotypes of adults and progeny remaining in the first well were scored as well as of the progeny in wells 1–3. Our screen was not ideal for detection of phenotypes visible only in adults (for example, egg-laying defective and progeny sterile); we will have missed some of these. Phenotypic analysis of lethality/sterility was carried out at $t = 24$ h and post-embryonic phenotypes were analysed by two independent observers at $t = 36, 48, 60$ and 72 h. Phenotypic classes were defined as follows. Embryonic lethal (Emb) reproducibly has 10–100% embryonic lethality; sterile (Ste) has a brood size of less than or equal to 10 (wild-type worms in these conditions typically give over 50); progeny sterile (Stp) has a brood size of less than or equal to 10 in the progeny of fed worms. Post-embryonic phenotypes require at least 10% of the analysed worms to display a given phenotype; phenotypic classes are given in the legend to Table 1. A full listing of phenotypes obtained is given in Supplementary Information Table 1; genes that we did not clone, and thus did not analyse, are given in Supplementary Information Table 3. Thus,

any GenePairs absent from both Supplementary Tables 1 and 3 was fed, and did not give a detectable mutant phenotype.

Bioinformatic analyses and categorization of genes

Analyses were carried out on GenePairs predictions rather than currently predicted genes as although gene predictions change, phenotypes will always match the GenePairs. About 95% of GenePairs genes have a one-to-one match with a currently predicted gene. Current gene predictions that are targeted for RNAi by the primer pairs were identified by comparing electronic PCR (ePCR) fragments (generated using the ePCR program (ftp.ncbi.nlm.nih.gov/pub/schuler/e-PCR)³¹ on the whole chromosome DNA files from the WS9 release of ACeDB (ftp.sanger.ac.uk/pub/wormbase) to gene predictions in ACeDB. To identify additional genes that might be targeted for RNAi by a particular clone we found those with an overlap of 200 base pairs or more with greater than 80% nucleotide identity with the predicted PCR product (asterisks in column 2 of Table 3 denote GenePairs that have such matches); however, it is not yet known what level of identity is required for RNAi.

To find *C. elegans* genes with conservation in other organisms, BlastP³² was carried out for each individual *C. elegans* gene on chromosome I against *S. cerevisiae*, *Drosophila melanogaster* and human sequences. The databases used were as follows: *C. elegans* (18,337 entries), *S. cerevisiae* (6,191 entries) and *D. melanogaster* (13,743 entries) downloaded on 1 June 2000 (<http://www.ebi.ac.uk/protome>); and *Homo sapiens* (35,723 entries, confirmed peptides) downloaded on 1 June 2000 (<http://www.ensembl.org>). NCBI-Blast2 was used (BLASTP 2.0.6) with the SEG filter, and the search space was set to 7,947,758.

We defined 'sequenced genes with a known phenotype' as being those with a named entry in ACeDB that also have a known phenotype entered in ACeDB, WormBase (<http://www.wormbase.org>) or the Proteome database (<http://www.proteome.com>). EST data were supplied by the Sanger Centre on 21 June 2000.

Predicted gene products were placed into functional classes by manual inspection, primarily using data from Proteome, InterPro and Blast analysis^{23,32}. The functional classes are: (1) DNA synthesis; (2) RNA synthesis and processing including general transcription machinery, splicing/processing, RNA binding and regulation of chromatin; (3) protein synthesis and proteolysis including translation, degradation and folding; (4) metabolism including energy production and intermediary metabolism; (5) cell-cycle and chromosome dynamics; (6) cell biology and cellular structure including cell junction/adhesion, cytoskeleton, ion channels, protein trafficking and vesicle regulation, and cell polarity; (7) gene-specific transcription; and (8) signal transduction including kinases, phosphatases and components of signal transduction pathways. The 'unknown' functional class contains either genes that have motifs about which there is insufficient information to assign a function, or genes with no significant matches in any organism.

Estimates of non-redundant genome size were done as follows. We detected 90.5% of genes known to give an embryonic lethal phenotype and 32.6% of genes known to give a post-embryonic phenotype. After screening 87.3% of the genes on chromosome I, we identified 226 Emb genes and 113 genes that only gave a post-embryonic RNAi phenotype (including steriles); adjusting for our efficiencies of detection, we estimate that on chromosome I, 286 genes should be required for viability and ~397 for post-embryonic processes. We screened 12.7% of the genome, and thus for the entire genome we expect 2,250 Emb genes and 3,130 total genes to have a post-embryonic phenotype.

Received 20 July; accepted 18 October 2000.

1. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
2. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
3. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
4. Ahringer, J. Turn to the worm! *Curr. Opin. Genet. Dev.* **7**, 410–415 (1997).
5. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
6. Tabara, H., Grishok, A. & Mello, C. C. RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282**, 430–431 (1998).
7. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
8. Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *C. elegans*. *Genome Biol.* (in the press).

9. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
10. Bowerman, B. & Shelton, C. A. Cell polarity in the early *Caenorhabditis elegans* embryo. *Curr. Opin. Genet. Dev.* **9**, 390–395 (1999).
11. Bargmann, C. I. Neurobiology of the *Caenorhabditis elegans* genome. *Science* **282**, 2028–2033 (1998).
12. Kim, J. G. & Hudson, L. D. Novel member of the zinc finger superfamily: A C2-HC finger that recognizes a glia-specific gene. *Mol. Cell Biol.* **12**, 5632–5639 (1992).
13. Armstrong, R. C., Kim, J. G. & Hudson, L. D. Expression of myelin transcription factor I (MyTI), a "zinc-finger" DNA-binding protein, in developing oligodendrocytes. *Glia* **14**, 303–321 (1995).
14. Kim, J. G. *et al.* Myelin transcription factor I (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system. *J. Neurosci. Res.* **50**, 272–290 (1997).
15. Strumpf, D. & Volk, T. Kakapo, a novel cytoskeletal-associated protein is essential for the restricted localization of the neuregulin-like factor, vein, at the muscle-tendon junction site. *J. Cell Biol.* **143**, 1259–1270 (1998).
16. Gregory, S. L. & Brown, N. H. kakapo, a gene required for adhesion between and within cell layers in *Drosophila*, encodes a large cytoskeletal linker protein related to plectin and dystrophin. *J. Cell Biol.* **143**, 1271–1282 (1998).
17. Lee, S., Harris, K. L., Whittington, P. M. & Kolodziej, P. A. *short stop* is allelic to *kakapo*, and encodes rod-like cytoskeletal-associated proteins required for axon extension. *J. Neurosci.* **20**, 1096–1108 (2000).
18. Prokop, A., Uhler, J., Roote, J. & Bate, M. The *kakapo* mutation affects terminal arborization and central dendritic sprouting of *Drosophila* motoneurons. *J. Cell Biol.* **143**, 1283–1294 (1998).
19. Burridge, K. & Connell, L. Talin: a cytoskeletal component concentrated in adhesion plaques and other sites of actin-membrane interaction. *Cell Motil.* **3**, 405–417 (1983).
20. Hodgkin, J. Sex, cell death, and the genome of *C. elegans*. *Cell* **98**, 277–280 (1999).
21. McNally, F. J. & Vale, R. D. Identification of katanin, an ATPase that severs and disassembles stable microtubules. *Cell* **75**, 419–429 (1993).
22. Srayko, M., Buster, D. W., Bazirgan, O. A., McNally, F. J. & Mains, P. E. MEI-1/MEI-2 katanin-like microtubule severing activity is required for *Caenorhabditis elegans* meiosis. *Genes Dev.* **14**, 1072–1084 (2000).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
24. Barnes, T. M., Kohara, Y., Coulson, A. & Hekimi, S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**, 159–179 (1995).
25. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 563–567 (1996).
26. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28**, 37–40 (2000).
27. Svetlov, V. V. & Cooper, T. G. Review: compilation and characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast* **11**, 1439–1484 (1995).
28. Hodgkin, J. & Herman, R. K. Changing styles in *C. elegans* genetics. *Trends Genet.* **14**, 352–357 (1998).
29. Chen, J. *et al.* Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics* **51**, 313–324 (1998).
30. Walhout, A. J. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**, 116–122 (2000).
31. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res.* **7**, 541–550 (1997).
32. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We would like to thank A. Fire and L. Timmons for providing protocols, feeding vectors and the HT115(DE3) bacterial strain; and R. Durbin, D. St. Johnston and B. Schumacher for critical reading of the manuscript. A.G.F. was supported by a US Army Breast Cancer Research Fellowship, R.S.K. by a Howard Hughes Medical Institute Predoctoral Fellowship, M.M.-C. by an EC-TMR Network Grant, P.Z. by a Wellcome Trust Studentship, M.S. by a Swiss National Science Foundation fellowship and J.A. by a Wellcome Trust Senior Research Fellowship.

Correspondence and requests for materials should be addressed to J.A. (e-mail: jaa@mole.bio.cam.ac.uk).

Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III

Pierre Gönczy^{*†‡}, Christophe Echeverri^{*†‡}, Karen Oegema^{*†}, Alan Coulson[§], Steven J. M. Jones^{||}, Richard R. Copley[†], John Duperon^{*}, Jeff Oegema^{*}, Michael Brehm^{*‡}, Etienne Cassin^{*}, Eva Hannak^{*}, Matthew Kirkham^{*}, Silke Pichler^{*†}, Kathrin Flohrs^{*}, Anoesjka Goessen^{*}, Sebastian Leidel^{*}, Anne-Marie Alleaume^{*‡}, Cécilie Martin^{*‡}, Nurhan Özlü^{*}, Peer Bork[†] & Anthony A. Hyman^{*†}

^{*} Max-Planck-Institute for Cell Biology and Genetics (MPI-CBG), Pfotenhauerstrasse 108, D-01307 Dresden, Germany

[†] European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, D-69117 Heidelberg, Germany

[§] The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

^{||} Genome Sequence Centre, British Columbia Cancer Research Centre, 600 West 10th Avenue, Vancouver, British Columbia, V5Z-4E6, Canada

[‡] Present addresses: Swiss Institute for Experimental Cancer Research (ISREC), CH-1066 Epalinges/Lausanne, Switzerland (P.G.); Cenix BioScience GmbH, Pfotenhauerstrasse 108, D-01307 Dresden, Germany (C.E., M.B., A.-M.A., C.M.).

Genome sequencing projects generate a wealth of information; however, the ultimate goal of such projects is to accelerate the identification of the biological function of genes. This creates a need for comprehensive studies to fill the gap between sequence and function. Here we report the results of a functional genomic screen to identify genes required for cell division in *Caenorhabditis elegans*. We inhibited the expression of ~96% of the ~2,300 predicted open reading frames on chromosome III using RNA-mediated interference (RNAi). By using an *in vivo* time-lapse differential interference contrast microscopy assay, we identified 133 genes (~6%) necessary for distinct cellular processes in early embryos. Our results indicate that these genes represent most of the genes on chromosome III that are required for proper cell division in *C. elegans* embryos. The complete data set, including sample time-lapse recordings, has been deposited in an open access database. We found that ~47% of the genes associated with a differential interference contrast phenotype have clear orthologues in other eukaryotes, indicating that this screen provides putative gene functions for other species as well.

Genome sequencing projects identify large numbers of genes in the search for biological function. Function can be tested directly in the organism of interest by using an appropriate assay or inferred on the basis of knowledge gained from the study of related genes in other organisms. However, a significant fraction of genes uncovered by sequencing projects are 'new' and cannot be rapidly assigned functions by either method. Thus, there is a need for large-scale studies to test directly the function of new genes, and validate gene functions inferred from studies of homologous proteins in other organisms. Large-scale functional genomic studies have already been successful in *S. cerevisiae* (ref. 1; <http://www.mips.biochem.mpg.de/proj/eurofan/index.html>; <http://genome-www.stanford.edu/cgi-bin/sgd/search>), but comparable analysis is still lacking in metazoans. Therefore, one of the challenges facing biologists today is to design high-throughput assays to assign cellular functions to genes emerging from metazoan sequencing projects.

Designing a large-scale screen for cell division genes

We sought to use a functional genomic approach in the early

C. elegans embryo to identify a large set of genes necessary for cell-division processes, for the following reasons. First, the genome is sequenced, offering the possibility to conduct a comprehensive analysis². Second, cell-division processes can be examined with high spatial and temporal resolution using time-lapse differential interference contrast (DIC) microscopy, allowing detection of even subtle deviations from the wild-type sequence of events³. Third, using RNA-mediated interference (RNAi), expression of a given gene in the embryo can be abolished in a sequence-specific manner by subjecting parental germ cells to corresponding double-stranded RNA (dsRNA)⁴. Notably, RNAi can be also applied to genes essential for cell division, as no mitoses occur between the time germ cells are subjected to dsRNA and fertilization. In contrast, although the targeted disruption of the ~6,200 open reading frames (ORFs) of

Table 2 Overall outcome of screen

Distinct genes tested	2,174
Total with phenotype	281 (~12.9%)
DIC phenotypes	133 (~6.1%)
DIC + EL	125*
DIC + L/A	5
DIC only	3†
Progeny phenotypes (only)	139 (~6.4%)
EL only	78
L/A only	61
L1/L2	33
L3/L4	15‡
Adult	13
FO sterile	9 (~0.4%)

EL, embryonic lethal; L/A, larval/adult phenotype; L1/L2, early larval defect; L3/L4, late larval defect; Adult, adult defect. See Methods, Table 3 and database (<http://mpi-web.embl-heidelberg.de/dbScreen/>) for more information.

*In some cases, embryonic lethality was partial only; progeny test not conducted for *par-2* (F58B6.3), but assumed to be embryonic lethal.

†For one of these dsRNAs (Y79H2A.11/Y75B8.36), the progeny test became embryonic lethal if conducted 36 h after injection (instead of the usual 24 h).

‡For one of these dsRNAs (ZK1236.3), larvae died as L3/L4.

Table 1 Positive controls

	Phenotype
Undiluted	13/13 (100%)
1:1 dilution	12/13 (~92%)
1:3 dilution	6/13 (~46%)
1:7 dilution	4/13 (~31%)

Double-stranded RNA corresponding to 13 genes whose phenotype in the early embryo is known from mutational analysis were generated and injected, either undiluted or diluted (1:1, 1:3, 1:7) with a mixture of 3–7 different dsRNA species corresponding to genes not associated with phenotypes. The resulting embryos were analysed by time-lapse DIC microscopy; phenotypes were scored as described in Methods. The fraction of dsRNAs giving rise to the expected phenotype is given for each dilution.

The following 13 genes were tested *cyk-1* (ref. 14), *let-99* (ref. 35), *mei-1* (ref. 36), *nmy-2* (ref. 37), *ooc-3* (ref. 38), *par-1* (ref. 39), *par-2* (ref. 18), *par-3* (ref. 17), *par-5* (K. Kemphues, personal communication), *par-6* (ref. 40), *zen-4* (refs 41, 42), *zyg-9* (ref. 43), *zyg-11* (ref. 44).

the *Saccharomyces cerevisiae* genome is yielding functional information at an unprecedented scale (<http://www.mips.biochem.mpg.de/proj/eurofan/index.html>; <http://genome-www.stanford.edu/cgi-bin/sgd/search>), the detailed cellular function of essential genes

cannot be examined in that paradigm. RNAi has been used previously as a reverse genetic tool to probe the function of individual genes, or undertake small-scale screening of a complementary DNA library⁵. We wanted to expand

Table 3 133 genes on chromosome III associated with DIC phenotypes

A. Meiotic divisions and early post-fertilization events

A1. Progress through meiotic divisions (11 genes)	
Male and female pronuclei do not become visible; embryos seem arrested during meiotic divisions.	
Gene	Brief identification
C02F5.9	proteasome component C5
C23G10.4	19S proteasome regulatory particle subunit
C30C11.2	26S proteasome regulatory subunit S3
F23F12.6	26S proteasome regulatory complex subunit p48A
F25B5.4	polyubiquitin (<i>ubq-1</i>)
F35G12.9	RING-finger containing protein (APC11-like)
F54C8.3	APC4 homologue (<i>emb-30</i>) *
F57B9.10	26S proteasome regulatory complex subunit p42B
K06H7.5	APC subunit 2
T05G5.3	Cdk1/Cdc2 kinase (<i>ncc-1</i>) *
ZK1010.1	ubiquitin (<i>ubq-2</i>)

A2. Fidelity of meiotic divisions (27 genes)	
Multiple female pronuclei; irregular cytoplasm; aberrant pseudocleavage stage; spindle unstable during anaphase; karyomeres in AB/P1; AB/P1 nuclei off-centre; often semi-sterile.	
B0336.10	60S ribosomal protein L17A
B0393.1	40S ribosomal protein SA
B0412.4	40S ribosomal protein S29
B0464.1	aspartyl-tRNA synthetase
C14B9.7	ribosomal protein L21
C16A3.9	40S ribosomal protein S13
C23G10.3	ribosomal protein S3
C27D11.1	translation initiation factor 3 subunit 10 (<i>egl-45</i>)
C54C6.1	60S ribosomal protein L37
F13B10.2	60S ribosomal protein L3
F26F4.10	arginyl tRNA synthetase
F37C12.4	ribosomal protein YL39
F37C12.9	ribosomal protein S14
F37C12.11	ribosomal protein S21
F53A3.3	40S ribosomal protein
F54E7.2	ribosomal protein S12
F56F3.5	ribosomal protein S3a
F57B9.3	eukaryotic initiation factor 4A
H06I04.4	ubiquitin-like ribosomal protein S27A fusion protein
R08D7.3	translation initiation factor eIF3 p66 subunit
R13A5.8	ribosomal protein L9
R74.1	leucyl-tRNA synthetase
R151.3	ribosomal protein ML16
T05G5.10	translation initiation factor eIF5A
T10F2.1	glycyl-tRNA synthetase
T20H4.3	prolyl-tRNA synthetase region
ZK652.4	60S ribosomal protein L35

A3. Entry into interphase (2 genes)	
Delay before entering interphase; vigorous cytoplasmic and cortical movements; aberrant number and/or position of pronuclei.	
F48E8.5	protein phosphatase 2A regulatory subunit
ZK520.4	cullin (<i>cul-2</i>) *

A4. Pseudocleavage stage (2 genes)	
Little/no cortical ruffling or pseudocleavage furrow.	
C34C12.3	serine/threonine protein phosphatase
Y49E10.19	homologies to anilin (actin-binding protein)

B. Nuclear appearance

B1. Pronuclear/nuclear appearance (5 genes)	
Pronuclei and nuclei in daughter blastomeres are not/poorly visible; spindle not/poorly visible; often failure in cytokinesis	
C29E4.3	Ran-GAP1
C38D4.3	weak homologies to calpastatin
F59A2.1	Ran-binding protein 2 (NUP 358)
K01G5.4	Ran
ZK328.5	NUP 98 (nucleoporin)

B2. Nuclear appearance (5 genes)	
Nuclei in daughter blastomeres are not/poorly visible	
C29E4.2	no clear homologies
F34D10.2	strong homologies to CDC45
R10E4.4	DNA replication licensing factor MCM5
Y55B1BR.3	contains chromo domain
ZK632.1	DNA replication licensing factor MCM6

C. Cell division processes in the early embryo

C1. Pronuclear migration (6 genes)	
Lack of male pronuclear migration; female pronuclear migration variable; sometimes multiple female pronuclei; no or small spindle.	
C05D11.3	putative ATP binding protein
C28H8.12	dynamitin (p50) (<i>dnc-2</i>) *
C36E8.5	beta-tubulin
K01G5.7	beta-tubulin
T03F6.5	LIS-1 homologue
T26A5.9	dynein 8kd light chain

C2. Spindle assembly (2 genes)	
Spindle is very small or not visible; karyomeres are generated	
F58A4.8	gamma-tubulin
H04J21.3	no clear homologies

C3. Fidelity of mitotic divisions (cross-eye phenotype) (3 genes)	
Daughter nuclei stay close to central cortex; usually karyomeres in daughter blastomeres.	
C02F5.1	homologies to coiled coil containing proteins
F35G12.8	SMC-1 condensin
F58A4.3	histone H3-like protein (homology to CENP-A)

C4. Fidelity of mitotic divisions (karyomeres only) (3 genes)	
Karyomeres in daughter blastomere AB and/or P1.	
F54C8.2	histone H3-like protein (homology to CENP-A)
R107.6	homologous to <i>Drosophila</i> orbit
Y43F4B.6	kinesin (KIF4-like)

C5. Anaphase spindle positioning (symmetric division) (4 genes)	
No posterior spindle displacement during anaphase; symmetric division.	
C38C10.4	no clear homologies outside <i>C. elegans</i>
F22B7.13	no clear homologies outside <i>C. elegans</i>
F54E7.3	PDZ containing protein PAR-3 (<i>par-3</i>) *
F58B6.3	RING-finger containing protein PAR-2 (<i>par-2</i>) *

C6. Cytokinesis (5 genes)	
Cleavage furrow not visible or regresses.	
B0464.5	serine/threonine kinase; similar to <i>S. pombe</i> DSK-1
C56G7.1	nonmuscle myosin regulatory light chain (<i>mlc-4</i>) *
F11H8.4	formin homology protein CYK-1 (<i>cyk-1</i>) *
K08E3.6	Rho family GTPase activating protein (<i>cyk-4</i>) *
T25C8.2	actin (<i>act-5</i>)

C7. P1 rotation (2 genes)	
No rotation of centrosome/nuclear complex in P1.	
F55H2.3	let-99 homologue
Y39E4B.1	RNAse L inhibitor

D. Pace of development

D1. General pace of development (overall slow) (12 genes)	
Slow overall pace of development (>30 min between pronuclear migration and AB division -compared to 18-22 min in wt).	
C29E4.8	adenylate kinase
C34E10.6	ATP synthase beta chain
F23H11.5	no clear homologies
F35G12.2	isocitrate dehydrogenase
F35G12.10	ATP synthase B chain
F54H12.1	aconitate hydratase
F56D2.1	mitochondrial processing protease enhancing protein
K04G7.4	NADH dehydrogenase
T07C4.7	succinate dehydrogenase cytochrome b chain (<i>mev-</i>)
T20G5.2	citrate synthase
T27E9.1	ADP/ATP mitochondrial carrier protein
ZK637.8	vacuolar H ⁺ -ATPase (TJ6/proton pump) (<i>unc-32</i>)

D2. Pace of development (8 genes)	
Slow between pseudocleavage stage and pronuclear envelope breakdown; P1 division delayed with respect to that of AB.	
C03C10.3	ribonucleotide reductase small subunit
F31E3.3	replication factor C complex protein
F44B9.7	replication factor C subunit 3
F58A4.4	DNA primase 49kd subunit
R01H10.1	DNA polymerase alpha/primase complex chain B
T23G5.1	ribonucleotide-disphosphate reductase large chain
T24C4.5	DNA primase subunit
Y47D3A.29	DNA polymerase alpha-subunit

E. Embryo appearance and morphology

E1. Osmotic integrity and other processes (16 genes)	
Embryos loose structural integrity upon dissection ; limited phenotypic analysis done in utero.	
B0336.2	ADP ribosylation factor 1 (<i>arf-1</i>)
B0361.10	glycosyl transferase group 1 protein
C07H6.5	RNA helicase (DEAD-box protein family)
C36A4.4	UDP-N-acetylglucosamin
D2045.1	no clear homologies
F08F8.2	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
F55H2.2	probable vacuolar ATP synthase subunit D
K10D2.6	NADPH-cytochrome P450
K12H4.4	signal peptidase subunit
PAR2.4	homologies to proteins in other metazoans
T05G5.7	no clear homologies outside <i>C. elegans</i>
T12A2.2	oligosaccharyl transferase STT3 subunit
ZK328.1	ubiquitin carboxyl-terminal hydrolase
ZK512.5	homologies to <i>Drosophila</i> and human proteins
ZK686.3	homologies to oligosaccharyl transferase 34 kd subunit
Y76A2B.1	coronin-like protein (<i>pod-1</i>) *

E2. Cytoplasmic appearance (sparse yolk granules) (4 genes)	
Markedly reduced density of yolk granules throughout embryo.	
C45G9.5	no clear homologies
T20G5.1	clathrin heavy chain
ZK1098.5	putative secretory protein (Bet3p-like)
Y37D8A.10	homologies to signal peptidase complex 25 kDa

E3. Cytoplasmic appearance (irregular) (3 genes)	
Uneven or irregular distribution of yolk granules.	
C03C10.1	casein kinase I
T20B12.1	homologies to O-linked GlcNAc transferase
Y49E10.15	small nuclear ribonucleoprotein E (snRNP E)

F. Unique phenotypes (13 genes)

Phenotypes associated with single genes: see comprehensive Table 3 in supplementary information for specifics.	
C07G2.3	T-complex protein 1, epsilon subunit (<i>cct-5</i>)
C14B9.4	polo-like kinase (<i>plk-1</i>) *
C18D11.5	homologies with transcription factors
F01F1.8	T-complex protein 1, zeta subunit (<i>cct-6</i>)
F10E9.8	no clear homologies
H38K22.2	homologies to proteins in other eukaryotes
K11D9.1	kinesin CeMCAK (KIF2/XKCM1-like)
R151.9	homologies to c-myc binding protein MM-1
T04A8.7	1,4-alpha-glucan branching enzyme
Y79H2A.11	doublecortin-related kinase
Y41C4A.10	Elongin B (regulator of RNA polymerase II)
Y49E10.1	26S proteasome regulatory subunit 8
Y56A3A.20	CCRA-associated factor 1

The gene name and a brief protein identification are given, as is a brief description of the phenotype characteristic of each class. Asterisk indicates that the phenotype for this gene was previously reported. Phenotypes represented by a single gene (two in the case of *cct-5* and *cct-6*) are grouped in 'unique phenotypes' for simplicity; specifics about each unique phenotype can be found in the Supplementary Information. A comprehensive version of Table 3 is available in the Supplementary Information. Additional details and time-lapse recordings can be found at <http://mpi-web.embl-heidelberg.de/dbScreen/>.

on these studies and apply RNAi on a genomic scale to identify genes required for cell division. We therefore sought to devise a screen that allows unbiased testing of each predicted ORF, not only of those represented in cDNA libraries. Here we report the design of such a screen, and its application to test the function of 2,232 (~96%) of the predicted ORFs on chromosome III of *C. elegans*.

Establishing an efficient screening paradigm

In brief, the screening strategy was established as follows. Primer pairs corresponding to individual ORFs were selected using a customized algorithm, with addition of T7 and T3 promoter sequences. Corresponding pieces of wild-type genomic DNA were PCR-amplified, separate T3 and T7 RNA transcription reactions performed and the resulting single-stranded RNAs (ssRNAs) annealed to generate dsRNA. After verifying their quality by gel electrophoresis, dsRNAs were micro-injected into wild-type hermaphrodites, and the resulting embryos were analysed 24 h later using time-lapse DIC microscopy. Single embryos were filmed from shortly after fertilization until the four-cell stage. This DIC assay constituted the core of the screen, and was optimized for detecting even minor deviation from the wild-type sequence of events. At the time of filming, three injected animals were also transferred to a fresh plate, which was scored 2 and 4 days later using low-magnification stereomicroscopy for the presence of progeny, and for any gross deviation from wild-type development.

To test whether this screening strategy would lead to the efficient identification of genes required in the early embryo, we injected individual dsRNAs corresponding to 13 genes characterized previously by mutational analysis to give rise to a phenotype visible by time-lapse DIC microscopy. As reported in Table 1, the expected phenotype was identified in all 13 cases. To determine the pool size optimal for large-scale studies, each of the 13 dsRNAs was mixed with 1, 3 or 7 volumes of dsRNA corresponding to genes not associated with phenotypes. Such dilutions substantially dampened the potency of RNAi (Table 1). Dampening was not observed when dsRNAs were diluted with either injection buffer or ssRNA (data not shown). These results suggest that there is a rapidly titratable step necessary for dsRNAs to generate RNAi. On the basis of these results, we chose a pool size of two to achieve higher throughput but still ensure identification of over 90% of the predicted genes required in the early embryo.

Pairs of randomly chosen dsRNAs were injected, and the resulting progeny were analysed with both DIC assay and progeny test. Pairs giving rise to a phenotype were split, and the two individual dsRNAs tested in turn. Each dsRNA associated with a phenotype was analysed further to ascertain that the phenotype did result from inactivation of the expected gene. This was necessary because of two potential caveats. First, individual dsRNAs can abolish expression of several genes closely related at the nucleotide level^{16,7}. Therefore, a BlastN homology search was performed with the PCR-amplified piece of DNA against the *C. elegans* genome. When this revealed sequences with more than 80% nucleotide identity over 200 base pairs (bp), or more than 90% over 100 bp, new primers were designed wherever possible to avoid regions of high sequence similarity. Second, mispredicted gene boundaries may result in some dsRNAs that also inactivate neighbouring genes. Therefore, dsRNAs whose neighbour was also associated with a phenotype were identified. When the neighbour was predicted to be less than 1 kb away from the amplified DNA, new primer pairs were designed wherever possible to increase that distance to more than 1 kb. In addition to these two steps, an independent dsRNA was generated for most genes associated with a DIC phenotype to verify the validity of the initial result (see Supplementary Information). Those genes whose RNAi phenotype was lost during this confirmation round were excluded from the data set. As a result, the DIC

phenotypes reported here almost certainly result from inactivation of the expected genes.

Analysis of ~96% of genes on chromosome III

We tested 2,232 of the 2,315 predicted ORFs (96.4%) on chromosome III and found that dsRNAs corresponding to 133 genes (6.1% of distinct genes tested) gave rise to a phenotype detectable in the DIC assay (Tables 2 and 3). dsRNAs corresponding to 139 genes gave rise to a phenotype detectable solely in the progeny test: 78 genes were associated with embryonic lethality, 48 with a larval phenotype; 13 with an adult phenotype; in addition, 9 dsRNAs resulted in sterility of the injected animals (Table 2; and Supplementary Information Table 4).

We compared these results to those obtained with classical genetics (see Supplementary Information). This showed that we had identified all seven chromosome III genes known to mutate to a phenotype recognizable by DIC microscopy in the early embryo, and the four genes whose role in the early embryo had been previously unravelled using RNAi^{5,8–18}. Moreover, we identified 9 out of 14 (~64%) loci known to mutate to phenotypes detectable later during embryogenesis, and 9 out of 31 (~29%) loci known to mutate to phenotypes in larvae or adults and that could have been easily recognized in the progeny test. These observations are compatible with a previous report that RNAi is less potent in silencing gene expression later during development⁴. Longer exposure to dsRNA, after injection or feeding, may help circumvent this limitation.

Notably, this analysis suggests that we have identified the large majority of the genes on chromosome III that are required for cellular processes scoreable by DIC microscopy in the early embryo. As we tested ~12% of the predicted ORFs in the genome, and assuming a random genome-wide distribution, it can be inferred that over 1,000 genes are essential for the first two cleavage divisions of *C. elegans*.

RNAi as an efficient gene function discovery tool

Detailed analysis of time-lapse recordings allowed us to place the 133 genes associated with a DIC phenotype into distinct phenotypic classes (Table 3). Owing to space limitations, only some classes are discussed here and illustrated in Fig. 1. The complete data set, along with time-lapse DIC recordings, can be found on an open-access database (<http://mpi-web.embl-heidelberg.de/dbScreen/>).

Of the 133 genes associated with a DIC phenotype, only 11 (8.2%) had been previously ascribed a function by direct experimentation in the early *C. elegans* embryo^{5,8–18}. Another 104 (78.2%) have homologues in other species, from which a function in *C. elegans* might potentially be derived. For the remaining 18 (13.5%), sequence information alone did not yield obvious clues to a possible function. Therefore, this screen represents an efficient discovery tool to assign gene function in the cognate cellular context. We illustrate this point with three phenotypic classes.

In the 'pronuclear/nuclear appearance' phenotypic class (B1 in Table 3), comprising five genes, pronuclei in the one-cell stage embryo, as well as nuclei in daughter blastomeres, are poorly or not visible (Fig. 1b). In addition, the spindle is poorly or not visible, and the cleavage furrow often regresses, probably as a consequence of the spindle defect. We infer that these genes are required for nuclear envelope assembly and an aspect of spindle formation. Notably, three of these genes encode the small GTPase Ran and its activating proteins Ran-GAP1 and Ran-BP2, which have a crucial role in nuclear envelope assembly and spindle formation in *Xenopus* egg extracts^{19–24}. Our work confirms that these molecules have a similar role *in vivo*. A fourth gene in this phenotypic class encodes the *C. elegans* orthologue of the nucleoporin NUP 98, and a fifth one a protein with weak homologies to calpastatin. An attractive hypothesis is that NUP 98 and a calpastatin-related protein may also

participate in nuclear envelope assembly in *C. elegans* and perhaps other metazoans.

In the 'pronuclear migration' phenotypic class (C1 in Table 3), comprising six genes, migration of pronuclei is affected (Fig. 1d). Two of the corresponding genes encode β -tubulin subunits, an

expected finding as pronuclear migration is a microtubule-dependent process²⁵. Two additional genes encode a dynein light chain and the dynactin component p50/dynamitin, which was also anticipated because cytoplasmic dynein function is required for this process in *C. elegans*⁹. In contrast, the involvement of the

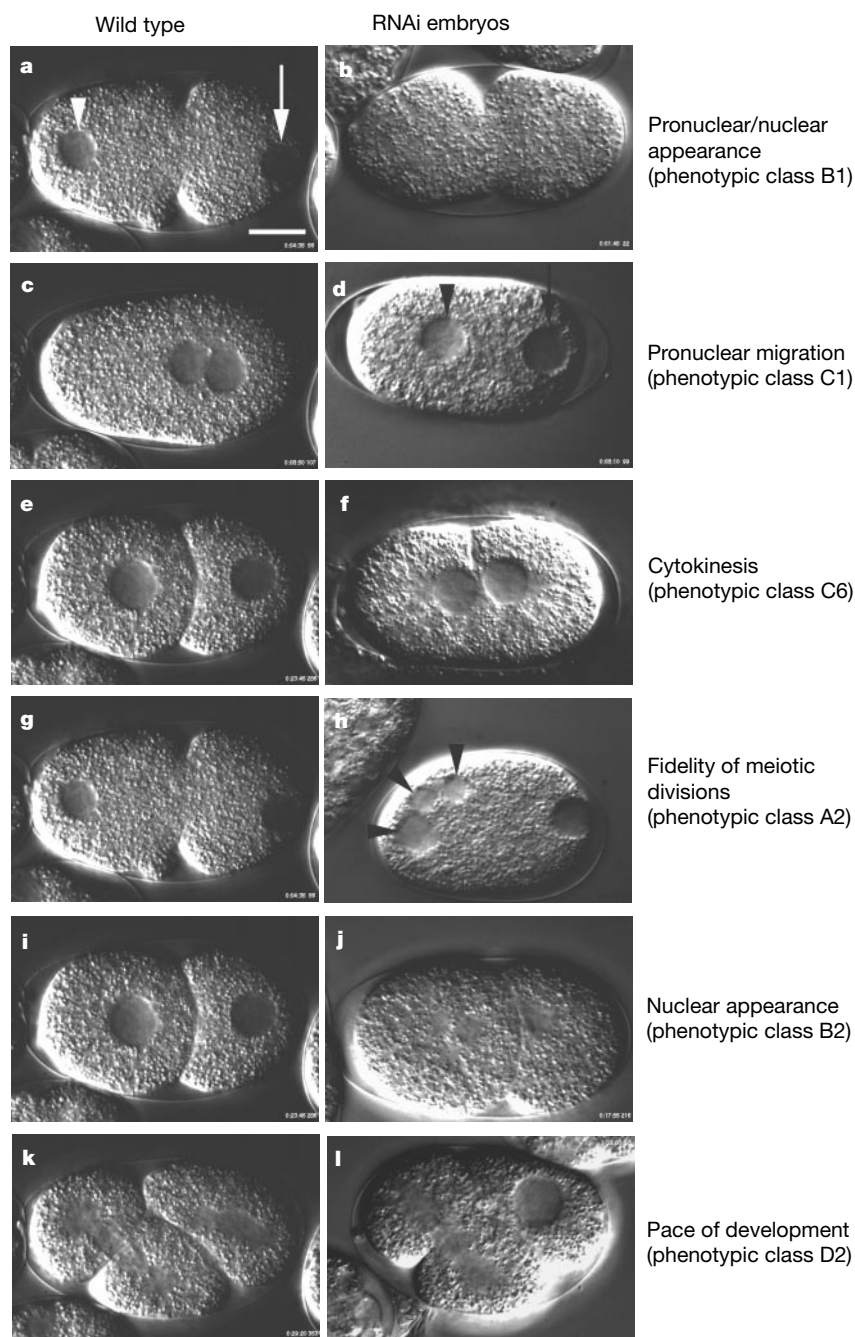


Figure 1 Single images taken from time-lapse DIC recordings of wild-type (left) and embryos representative of six distinct phenotypic classes (right). Anterior is to the left and posterior to the right. All panels at same magnification; scale bar, 10 μ m.

a, b, Pseudocleavage stage. In wild type (**a**), both male (arrow) and female (arrowhead) pronuclei are visible as clear discs excluding yolk granules. In contrast, in embryos of the pronuclear/nuclear appearance class (**b**; C29E4.3, RanGap1 shown here), pronuclei are not visible. Note presence of pseudocleavage furrow in both cases. **c, d**, Pronuclear migration stage. In wild type (**c**), the two pronuclei have met in the posterior half of the embryo. In embryos of the pronuclear migration class (**d**; T03G6.5, LIS-1 shown here), the male pronucleus (arrow) is still in the vicinity of the posterior cortex, the female pronucleus (arrowhead) is in the anterior of the embryo. **e–f**, Two-cell stage. In wild-type (**e**), blastomeres of the two-cell stage embryo have been clearly separated by the cleavage

furrow. In embryos of the cytokinesis class (**f**; B0464.5, serine/threonine kinase shown here), the cleavage furrow regresses, and daughter nuclei rejoin in a single cell. **g, h**, Pseudocleavage stage. In wild-type (**g**), there is a single female pronucleus. In embryos of the fidelity of meiotic divisions class (**h**; C16A3.9, 40S ribosomal protein S13 shown here), there are several female pronuclei (arrowheads). Note absence of pseudocleavage furrow. **i, j**, Two-cell stage. In wild-type (**i**), nuclei in daughter blastomeres are visible as clear discs excluding yolk granules. In embryos of the nuclear appearance class (**j**; F34D10.2, CDC45-like shown here), nuclei are not visible. **k, l**, About 20 min after pronuclear migration. In wild type (**k**), the anterior blastomere (AB) is completing division, whereas the posterior blastomere (P₁), follows 2–3 min later; note that the nuclear envelope in P₁ has already broken down. In embryos of the pace of development class (**l**; T24C4.5, DNA primase shown here), division of P₁ is delayed with respect to that of AB; note intact nuclear envelope in P₁.

remaining two genes was not documented previously in nematodes. We found that a putative ATP-binding protein and the worm homologue of *lis-1*, the gene mutated in Müller–Dieker lissencephaly, are also required for pronuclear migration. *lis-1* homologues act in concert with dynein to mediate nuclear localization in both *Aspergillus* and *Drosophila*^{26,27}, and our finding indicates that the two components may also function together in *C. elegans*.

In the ‘cytokinesis’ phenotypic class (C6 in Table 3), comprising five genes, cleavage is defective, resulting in the rejoining of daughter blastomeres (Fig. 1f). Three of these genes, *cyk-1*, *cyk-4* and *mlc-4*, which encode a formin¹⁴, a Rho GAP¹⁵ and a non-muscle myosin regulatory light chain⁵, respectively, have been initially identified by mutational analysis. This confirms that an RNAi-based functional genomic approach will provide a rich source for the molecular identification of loci defined by classical genetics.

Functional groups of genes required for cellular processes

Our RNAi-based functional genomic screen has also allowed us to reach unexpected conclusions about the involvement of particular biochemical pathways in specific cellular process in the early *C. elegans* embryo. We illustrate this point with three phenotypic classes.

In the ‘fidelity of meiotic divisions’ class (A2 in Table 3), comprising 27 genes, embryos often display several female pronuclei, indicative of defects during the female meiotic divisions (Fig. 1h). Additional phenotypic manifestations include an unstable spindle during anaphase and generation of karyomeres in daughter blastomeres. Unexpectedly, all 27 genes in this class encode components of the translation machinery, suggesting that efficient protein synthesis is required for proper execution of meiotic divisions in *C. elegans*. Whether this reflects a requirement for overall protein synthesis, or a requirement for the synthesis of a specific subset of proteins needed for the meiotic divisions remains to be determined.

In the ‘nuclear appearance’ phenotypic class (B2 in Table 3), comprising five genes, although pronuclei are visible as in wild-type, nuclei in daughter blastomeres are poorly or not visible (Fig. 1j). We infer that these genes are required specifically for the reassembly of the nucleus after mitosis. Three of these genes encode *C. elegans* orthologues of the DNA replication licensing factors CDC45, CDC46 and MCM2/3, which are required both for DNA replication and for restricting its occurrence to once per cell cycle²⁸. These findings suggest that there may possibly be a mechanism that monitors the presence of DNA licensing factors before nuclear

reassembly can be triggered.

In the ‘pace of development’ phenotypic class (D2 in Table 3), comprising eight genes, events are delayed in the one-cell stage between pronuclear meeting and pronuclear envelope breakdown, and in the two-cell stage most strikingly in the P₁ blastomere (Fig. 1l). Notably, six of these eight genes encode components of the DNA replication machinery, whereas the remaining two encode polypeptides of ribonucleotide-diphosphate reductase, which is required to produce deoxyribonucleotides. This indicates that the observed delays may be due to defects in DNA replication. This implicates the existence of an hitherto unsuspected DNA replication checkpoint in the early *C. elegans* embryo. This checkpoint mechanism may act preferentially in the P lineage, as events appear less delayed in the AB blastomere of the two-cell stage embryo.

Toward comprehensive metazoan functional genomics

We wanted to investigate whether genes identified in this screen are likely to have similar functions in other organisms. As an indirect way to address this issue, we determined whether genes associated with DIC phenotypes were more likely to be conserved across evolution. We classified all tested proteins into one of four groups based on BlastP analysis: (1) nematode-specific; (2) metazoan-specific; (3) eukaryotic-specific; (4) multi-kingdom. We found a distinct enrichment of evolutionarily conserved genes among those associated with DIC phenotypes (Fig. 2a). Most strikingly, although *C. elegans* genes that have orthologues in both *Drosophila* and *S. cerevisiae* represent only 12.9% of genes tested, they comprise 47.3% of those associated with DIC phenotypes (Fig. 2b). Therefore, genes required for proper division of the early *C. elegans* embryo tend to have close relatives in other species, suggesting that a number of the functional assignments made here may be transferred to other eukaryotes.

Our work shows that RNAi is an efficient reverse genetic tool to undertake a large-scale functional genomic analysis. Detailed analysis using time-lapse DIC microscopy has allowed us to identify ‘phenotypic signatures’ characteristic of functional groups of genes, and thus establish a unique functional database of cell-division proteins. By continuing this analysis on a genome-wide basis, by using RNAi to probe other processes with different assays, and by comparing the data with functional information emerging from other organisms, a significant fraction of metazoan genes may be assigned functions. □

Methods

Selection of primer pairs and number of genes tested

Primer pairs were selected using a customized program to amplify the shortest possible region containing > 500 bp or >90% coding sequence, with preference given to experimentally confirmed exons and regions covered by expressed sequence tags (ESTs) (see Supplementary Information).

We tested 2,232/2,315 chromosome III ORFs (Wormpep20) (see list at <http://mpi-web.embl-heidelberg.de/dbScreen/>). These correspond to 2,174 distinct genes because (1) 48 genes have two or more alternatively spliced variants; (2) F55H2.3/F55H2.4, K06H7.5/K06H7.6, Y79H2A.11/Y75B8A.36 each correspond to a single gene (Table 3); and (3) one gene associated with a phenotype (Y47D3A.29) was not present in Wormpep20.

Generation of dsRNA

All steps were done in 96-well plates. PCR reactions (50 µl) were performed using 0.8 µM primers and ~0.1 µg wild-type template genomic DNA. PCR products were ethanol-precipitated and resuspended in 7.0 µl TE buffer; 1.0 µl was used for separate 5 µl transcription reactions using T3 and T7 RNA polymerases (Ambion). The reactions were diluted to 50 µl with RNase-free water, combined, and then the mixed RNA was purified (Qiagen) into 130 µl RNase-free H₂O; 50 µl of this was mixed with 10 µl 6× injection buffer²⁹, and the RNA annealed by heating for 10 min at 68 °C, and 30 min at 37 °C. Concentrations of dsRNAs were 0.1–0.3 µg µl⁻¹. Double stranding was assessed by scoring shifts in mobility with respect to ssRNA on 1% agarose gels.

Injections and time-lapse DIC microscopy assay

Pairs of dsRNAs were injected bilaterally into the gonads of adult wild-type hermaphrodites²⁹, which were left at 20 °C for 24 h. Embryos were then removed and

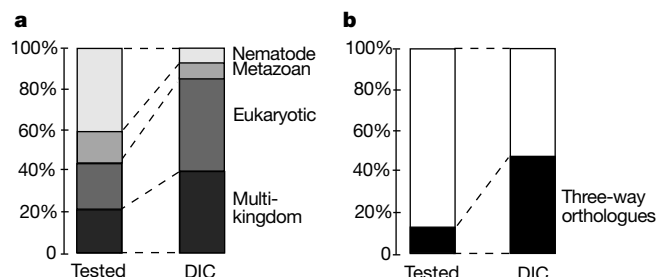


Figure 2 Distribution of predicted proteins according to homology and orthology relationships. Left, 2,233 genes tested (2,232 from Wormpep20 plus Y47D3A.29); right, 133 genes associated with DIC phenotypes. **a**, Homology. BlastP analysis was conducted to place predicted proteins into one of the following categories (from top to bottom, in increasing shadings): nematode-specific (39.9% in tested set, 7.5% in DIC set); metazoan-specific (16.3% in tested set, 7.5% in DIC set); eukaryotic-specific (23% in tested set, 45.1% in DIC set); multi-kingdom (20.7% in tested set, 39.8% in DIC set). **b**, Three-way (*C. elegans*, *Drosophila*, *S. cerevisiae*) reciprocal BlastP analyses were conducted to determine whether predicted proteins had clear orthologues in other eukaryotes (black shading: 12.9% in tested set; 47.3% in DIC set).

analysed for potential defects in cell-division processes, capturing 1 image every 5 s (ref. 3). Details about cellular processes scored are given in Supplementary Information. A minimum of three embryos from three different worms were filmed from shortly after fertilization until the four-cell stage. Embryos from two additional worms were also inspected and still images taken. Embryos resulting from injection of some dsRNAs lost structural integrity on dissection, and were examined *in utero*³.

Pairs of dsRNAs were split if one or more embryo displayed a phenotype. A phenotype was attributed to a given gene when it was observed in at least two embryos examined after injection of individual dsRNA.

Progeny test

Three animals were transferred to a fresh plate 24 h after injection, and left at 20 °C. Two days later, the plate was inspected with a stereomicroscope (×20–40 magnification) for the presence of eggs, F1 larvae and their developmental stage (normally L2–L4). Two days after that, the plate was inspected for the presence of F1 adults (normally >100), their overall body morphology and the presence of F2 progeny. Whenever the progeny test was conducted more than once with a given dsRNA and gave rise to different results, the strongest phenotype was retained.

Double-stranded RNAs that gave rise to eggs and 0–10 F1 larvae were deemed embryonic lethal. dsRNAs that gave rise to a phenotype detectable in larvae fell into two groups: those in which the progeny on day four resembled L1 or L2 ('early larval defect'); and those in which the progeny on day four resembled L3 or L4 ('late larval defect'). Partially penetrant embryonic lethality and subtle developmental defects were not scored in this analysis. Moreover, dsRNAs that gave rise to defects in less than 5% of the adult progeny were not considered as being associated with a phenotype.

Bioinformatic analysis

Homologues. *C. elegans* protein sequences were searched against the sp_nrd database, a non-redundant composite of SWISS-PROT and TrEMBL³⁰, using BlastP³¹ and masking compositionally biased regions with the SEG filter. Significantly scoring sequences (E-value < 10⁻⁶) were retrieved, and species and phyla information extracted from the sequence entry file. On the basis of these results, proteins were broken down into the following categories: (1) nematode specific (no significant homologue present outside nematodes); (2) metazoan specific (significant homologues present in other metazoans, but not outside metazoans); (3) eukaryotic specific (significant homologues present among unicellular eukaryotes, but not outside eukaryotes); (4) multi-kingdom (significant homologues present in both eukaryotes and prokaryotes). Although this procedure will not recognize very distant homologues, it will identify proteins that are well-conserved across kingdoms. Varying the E-value to <10⁻³ did not lead to major changes in the outcome of the analysis.

Orthologues. Using BlastP³¹, we identified proteins from *C. elegans*, *D. melanogaster* and *S. cerevisiae* that were all each other's best match in the respective genomes³². That two proteins in different genomes are each others host match is commonly used as an operational definition of orthology^{33,34}, although it is recognized that full reconstructions of phylogeny are required to resolve complex evolutionary histories. *C. elegans* protein data were obtained by embedding in Wormpep20 the protein sequences tested by RNAi. *Drosophila* protein sequence data were obtained from the Celera genome CD-ROM, *S. cerevisiae* protein data from the NCBI.

Received 17 July; accepted 17 October 2000.

- Ross-Macdonald, P. et al. Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Gönczy, P. et al. Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *J. Cell Biol.* **144**, 927–946 (1999).
- Fire, A. et al. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Shelton, C. A., Carter, J. C., Ellis, G. C. & Bowerman, B. The nonmuscle myosin regulatory light chain gene *mlc-4* is required for cytokinesis, anterior-posterior polarity, and body morphology during *Caenorhabditis elegans* embryogenesis. *J. Cell Biol.* **146**, 439–451 (1999).
- Zhang, B. et al. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* **390**, 477–484 (1997).
- Cox, D. N. et al. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
- Skop, A. R. & White, J. G. The dynactin complex is required for cleavage plane specification in early *Caenorhabditis elegans* embryos. *Curr. Biol.* **8**, 1110–1116 (1998).
- Gönczy, P., Pichler, S., Kirkham, M. & Hyman, A. A. Cytoplasmic dynein is required for distinct aspects of MTOC positioning, including centrosome separation, in the one cell stage *Caenorhabditis elegans* embryo. *J. Cell Biol.* **147**, 135–150 (1999).
- Chase, D. et al. The polo-like kinase PLK-1 is required for nuclear envelope breakdown and the completion of meiosis in *Caenorhabditis elegans*. *Genesis* **26**, 26–41 (2000).
- Boxem, M., Srinivasan, D. G. & van den Heuvel, S. The *Caenorhabditis elegans* gene *ncc-1* encodes a cdc2-related kinase required for M phase in meiotic and mitotic cell divisions, but not for S phase. *Development* **126**, 2227–2239 (1999).
- Feng, H. et al. CUL-2 is required for the G1-to-S-phase transition and mitotic chromosome condensation in *Caenorhabditis elegans*. *Nature Cell Biol.* **1**, 486–492 (1999).
- Rappleye, C. A., Paredes, A. R., Smith, C. W., McDonald, K. L. & Aroian, R. V. The coronin-like protein POD-1 is required for anterior-posterior axis formation and cellular architecture in the nematode *Caenorhabditis elegans*. *Genes Dev.* **13**, 2838–2851 (1999).

- Swan, K. A. et al. *cyk-1*: a *C. elegans* FH gene required for a late step in embryonic cytokinesis. *J. Cell Sci.* **111**, 2017–2027 (1998).
- Jantsch-Plunger, V. et al. CYK-4. A rho family gtpase activating protein (gap) required for central spindle formation and cytokinesis. *J. Cell Biol.* **149**, 1391–1404 (2000).
- Furuta, T. et al. EMB-30: an APC4 homologue required for metaphase-to-anaphase transitions during meiosis and mitosis in *Caenorhabditis elegans*. *Mol. Biol. Cell* **11**, 1401–1419 (2000).
- Etemad-Moghadam, B., Guo, S. & Kemphues, K. J. Asymmetrically distributed PAR-3 protein contributes to cell polarity and spindle alignment in early *C. elegans* embryos. *Cell* **83**, 743–752 (1995).
- Levitán, D. J., Boyd, L., Mello, C. C., Kemphues, K. J. & Stinchcomb, D. T. *par-2*, a gene required for blastomere asymmetry in *Caenorhabditis elegans*, encodes zinc-finger and ATP-binding motifs. *Proc. Natl Acad. Sci. USA* **91**, 6108–6112 (1994).
- Nicolas, F. et al. *Xenopus* Ran-binding protein 1: molecular interactions and effects on nuclear assembly in *Xenopus* egg extracts. *J. Cell Sci.* **110**, 3019–3030 (1997).
- Carazo-Salas, R. E. et al. Generation of GTP-bound Ran by RCC1 is required for chromatin-induced mitotic spindle formation. *Nature* **400**, 178–181 (1999).
- Kalab, P., Pu, R. T. & Dasso, M. The ran GTPase regulates mitotic spindle assembly. *Curr. Biol.* **9**, 481–484 (1999).
- Wilde, A. & Zheng, Y. Stimulation of microtubule aster formation and spindle assembly by the small GTPase Ran. *Science* **284**, 1359–1362 (1999).
- Hetzler, M., Bilbao-Cortes, D., Walther, T. C., Gruss, O. J. & Mattaj, I. W. GTP hydrolysis by Ran is required for nuclear envelope assembly. *Mol. Cell* **5**, 1013–1024 (2000).
- Zhang, C. & Clarke, P. R. Chromatin-independent nuclear envelope assembly induced by Ran GTPase in *Xenopus* egg extracts. *Science* **288**, 1429–1432 (2000).
- Strome, S. & Wood, W. B. Generation of asymmetry and segregation of germ-line granules in early *C. elegans* embryos. *Cell* **35**, 15–25 (1983).
- Swan, A., Nguyen, T. & Suter, B. *Drosophila* Lissencephaly-1 functions with Bic-D and dynein in oocyte determination and nuclear positioning. *Nature Cell Biol.* **1**, 444–449 (1999).
- Willins, D. A., Liu, B., Xiang, X. & Morris, N. R. Mutations in the heavy chain of cytoplasmic dynein suppress the nudF nuclear migration mutation of *Aspergillus nidulans*. *Mol. Gen. Genet.* **255**, 194–200 (1997).
- Tye, B. K. MCM proteins in DNA replication. *Annu. Rev. Biochem.* **68**, 649–686 (1999).
- Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
- Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
- Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Tatusov, R. L., Koonin, E. V. & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
- Chervitz, S. A. et al. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**, 2022–2028 (1998).
- Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. *Nature Genet.* **21**, 108–110 (1999).
- Rose, L. S. & Kemphues, K. The *let-99* gene is required for proper spindle orientation during cleavage of the *C. elegans* embryo. *Development* **125**, 1337–1346 (1998).
- Clark-Maguire, S. & Mains, P. E. *mei-1*, a gene required for meiotic spindle formation in *Caenorhabditis elegans*, is a member of a family of ATPases. *Genetics* **136**, 533–546 (1994).
- Guo, S. & Kemphues, K. J. A non-muscle myosin required for embryonic polarity in *Caenorhabditis elegans*. *Nature* **382**, 455–458 (1996).
- Pichler, S. et al. OOC-3, a novel putative transmembrane protein required for establishment of cortical domains and spindle orientation in the P₁ blastomere of *C. elegans* embryos. *Development* **127**, 2063–2073 (2000).
- Guo, S. & Kemphues, K. J. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**, 611–620 (1995).
- Hung, T. J. & Kemphues, K. J. PAR-6 is a conserved PDZ domain-containing protein that colocalizes with PAR-3 in *Caenorhabditis elegans* embryos. *Development* **126**, 127–135 (1999).
- Raich, W. B., Moran, A. N., Rothman, J. H. & Hardin, J. Cytoplasmic dynein and midzone microtubule organization in *Caenorhabditis elegans* require the kinesin-like protein ZEN-4. *Mol. Biol. Cell* **9**, 2037–2049 (1998).
- Powers, J., Bossinger, O., Rose, D., Strome, S. & Saxton, W. A nematode kinesin required for cleavage furrow advancement. *Curr. Biol.* **8**, 1133–1136 (1998).
- Matthews, L. R., Carter, P., Thierry, M. D. & Kemphues, K. ZYG-9, a *Caenorhabditis elegans* protein required for microtubule organization and function, is a component of meiotic and mitotic spindle poles. *J. Cell Biol.* **141**, 1159–1168 (1998).
- Carter, P. W., Roos, J. M. & Kemphues, K. J. Molecular analysis of *zyg-11*, a maternal-effect gene required for early embryogenesis of *Caenorhabditis elegans*. *Mol. Gen. Genet.* **221**, 72–80 (1990).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

For help in improving the manuscript, we thank A. Desai, A. Ephrussi, S. Grill, M. Labouesse, I. Mattaj and B. Sönnichsen. Work in the Hyman laboratory is supported by the EMBL (European Molecular Biology Laboratory) and the MPI (Max-Planck-Institute). P.G. was supported by a fellowship from the Swiss National Science Foundation. A.C. is supported by the UK Medical Research Council.

Correspondence and requests for materials should be addressed to P.G. (e-mail: Pierre.Gönczy@isrec.unil.ch) or A.A.H. (e-mail: hyman@EMBL-Heidelberg.de). The complete data set, including sample time-lapse recordings, has been deposited in an open-access database (<http://mpi-web.embl-heidelberg.de/dbScreen/>).

Mean-field cluster model for the critical behaviour of ferromagnets

Ralph V. Chamberlin

Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287-1504, USA

Two separate theories are often used to characterize the paramagnetic properties of ferromagnetic materials. At temperatures T well above the Curie temperature, T_C (where the transition from paramagnetic to ferromagnetic behaviour occurs), classical mean-field theory¹ yields the Curie–Weiss law for the magnetic susceptibility: $\chi(T) \propto 1/(T - \Theta)$, where Θ is the Weiss constant. Close to T_C , however, the standard mean-field approach breaks down so that better agreement with experimental data is provided by critical scaling theory^{2,3}: $\chi(T) \propto 1/(T - T_C)^\gamma$, where γ is a scaling exponent. But there is no known model capable of predicting the measured values of γ nor its variation among different substances⁴. Here I use a mean-field cluster model⁵ based on finite-size thermostatics^{6,7} to extend the range of mean-field theory, thereby eliminating the need for a separate scaling regime. The mean-field approximation is justified by using a kinetic-energy term to maintain the microcanonical ensemble⁸. The model reproduces the Curie–Weiss law at high temperatures, but the classical Weiss transition at $T_C = \Theta$ is suppressed by finite-size effects. Instead, the fraction of clusters with a specific amount of order diverges at T_C , yielding a transition that is mathematically similar to Bose–Einstein condensation. At all temperatures above T_C , the model matches the measured magnetic susceptibilities of crystalline EuO, Gd, Co and Ni, thus providing a unified picture for both the critical-scaling and Curie–Weiss regimes.

In 1873 van der Waals presented the first viable model for a phase transition⁹. His model involved coupling the local degrees of freedom to the average field arising from the rest of the sample. This concept, which is the basis of mean-field theory, remains the most versatile approach for characterizing the internal response of real substances such as dense gases, fluids and solids. In the 1940s, however, Onsager¹⁰ proved that the two-dimensional Ising model does not obey the standard mean-field predictions of a phase transition at $T_C = \Theta$ with $\gamma = 1$, and Guggenheim¹¹ showed that measured critical exponents differ systematically from classical values. Considerable insight into non-classical scaling was obtained in the 1960s and 70s, and modern theoretical techniques yield accurate critical exponents for most models with homogeneous interactions. Although basic scaling has provided a useful description over a range of temperatures near the transition, and corrections to scaling^{12,13} can extend this range, mean-field theory is still best for data at high temperatures.

Most previous cluster models^{14,15} involved evaluating the configurations of a single cluster, then coupling it to the mean field from the rest of the sample. However, as the correlation length diverges near T_C , very large clusters are necessary for accurate calculations, so that there is little or no computational advantage over evaluating the entire system. Here I consider clusters with unrestricted sizes, but use a mean-field expression for the interactions within each cluster, and ignore the direct interactions between clusters. The mean-field result greatly simplifies the mathematics, while the extra degrees of freedom from unrestricted sizes give the correct critical behaviour.

My model is equivalent to the original Weiss theory applied to finite clusters, so I begin with a brief review of the accepted description of ferromagnets at high temperatures. Let σ_i represent the alignment of the i th particle, so that in the Ising model σ_i may be

‘up’ (+1) or ‘down’ (−1). Although real spins in most materials have more than two available states, for simplicity I approximate these multiple degrees of freedom by multiple Ising-like quasiparticles, or ‘particles’ for short. For a cluster of m particles, with interaction strength $\epsilon_0/2z$ between each particle and its z nearest neighbours, the interaction energy per particle may be written¹⁵ $V_m(l, l_{++}) = -(\epsilon_0/4)(8l_{++}/zm - 4l/m + 1)$, where l is the number of ‘up’ particles and l_{++} is the number of pairs of interacting ‘up’ particles. Most of the difficulty in evaluating $V_m(l, l_{++})$ comes for the pair-interaction term l_{++} . In the mean-field approximation, this difficulty is avoided by using the average value $\bar{l}_{++}/m \approx (z/2)(l/m)^2$, leaving $\bar{V}_m(l) \approx -(\epsilon_0/4)(2l/m - 1)^2$. Although local correlations are known to cause significant deviations from infinite-ranged models near T_C , here I show that these correlations can be attributed to a similar mean-field model applied to finite clusters.

Starting with an isolated cluster of size m , I make two basic assumptions to justify the mean-field result. (1) Net angular momentum is conserved; thus changes in l can occur only through contact with neighbouring clusters, whereas spin-conserving changes in the local configuration (l_{++}) occur internally. (2) Net energy per particle ($\epsilon_m(l)$) is conserved by a balance between potential and kinetic energies; thus configurations with low potential energy have high kinetic energy and vice versa, so that the virial theorem for any realistic potential near a stable equilibrium yields $\epsilon_m(l) = 2\bar{V}_m(l)$. In the Ising model, energy-carrying ‘demons’ have been used as a reservoir of kinetic energy to maintain the microcanonical

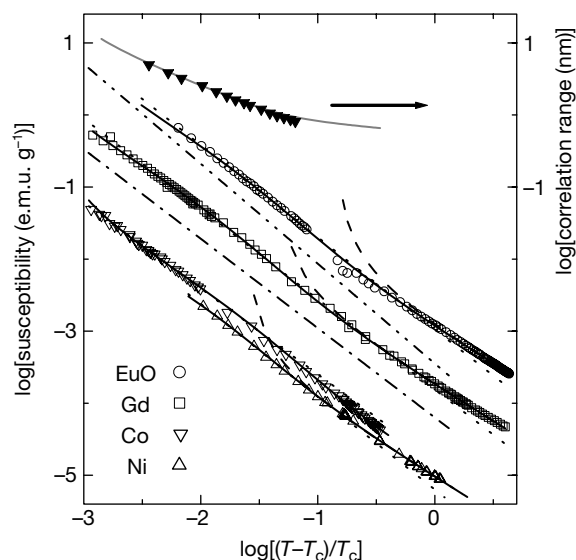


Figure 1 Scaling plot of the magnetic susceptibilities of crystalline europium oxide^{23,24}, gadolinium^{25–27}, cobalt^{28,29} and nickel^{30,31}. Each dashed curve is a best fit to $\chi(T) \propto 1/(T - \Theta)$ at high temperatures, giving the Weiss constant Θ . Note that Co has not been measured to high enough temperatures to show clear Curie–Weiss behaviour. Each dotted line is a best fit to $\chi(T) \propto 1/(T - T_C)^\gamma$ near T_C , giving the critical exponent γ . None of the substances has the Heisenberg-model slope $\gamma = 1.387$ (dot-dot-dashed line), not even EuO, which is among the most highly localized and isotropic spin systems known. Although Gd and Co show a range where their slopes are similar to the standard Ising value $\gamma = 1.240$ (dot-dashed line), close inspection reveals their slight sigmoidal curvature, with slopes very near to T_C that are somewhat less than the Ising model. Because the measured critical exponents do not correspond to any known homogeneous model, the combination of empirical scaling plus Curie–Weiss law requires five adjustable parameters. In contrast, the single expression equation (2) (solid black curves) gives good agreement for the entire paramagnetic phase, without having to introduce a separate transition temperature and amplitude prefactor for the scaling regime. The uppermost plots are of the magnetic correlation range (ξ) of Co, as measured by small-angle neutron scattering²² (filled triangles), and as deduced from the average number of particles in an aggregate, $(\overline{mn})^{1/3}$, normalized to give one atomic radius as $T \rightarrow \infty$ (solid grey curve).

ensemble⁸. (Note that an energy offset necessary to give $\epsilon_m(0) = 0$ has been neglected, but this offset is irrelevant for the changes in $\epsilon_m(l)$ that govern the thermal properties of the cluster.) Using Gibbs' fundamental postulates of statistical mechanics—that all allowed states are equally likely and that the equilibrium of a finite system is found from its time-averaged properties—the net energy per particle is $\epsilon_m(l) = -(\epsilon_0/2)[4l(l-1)/m(m-1) - 4l/m + 1]$.

One way to demonstrate the validity of $\epsilon_m(l)$ is to examine the interaction energies of a specific example. Consider four spins ($m = 4$), with two 'up' spins ($l = 2$), arranged in a square so that each spin interacts with its two nearest neighbours ($z = 2$). One set of four configurations has the aligned spins at adjacent corners of the square giving $V_4(2, 1) = 0$, while another set of two configurations has the aligned spins at diagonal corners giving $V_4(2, 0) = +\epsilon_0/4$. Because all six configurations are equally likely, the equilibrium potential energy per particle is $\bar{V}_4(2) = [4(0) + 2(\epsilon_0/4)]/6 = \epsilon_0/12$. Thus, even for a small cluster with only two nearest neighbours the net energy is in full agreement with $\epsilon_4(2) = \epsilon_0/6$; and for large clusters when $(l-1)/(m-1) \approx l/m$, this energy approaches the standard mean-field result. Indeed, except for a surface-tension term which can be neglected for large clusters, $\epsilon_m(l)$ is the exact equilibrium energy for the Ising interaction plus kinetic energy in the appropriate microcanonical ensemble.

Now let the cluster interact with an ensemble of other clusters. The exchange of angular momentum allows all possible values of l , the exchange of particles allows all possible values of m , and the exchange of energy defines the temperature T . Using the binomial coefficient for the degeneracy of each state, the partition function becomes:

$$\Gamma = \sum_{m=2}^{\infty} \sum_{l=0}^m \frac{m!}{(m-l)!l!} \exp\{-m[\epsilon_m(l, H) - \mu]/k_B T\} \quad (1)$$

Here μ is the chemical potential, $\epsilon_m(l, H) = \epsilon_m(l) + M_0 H(2l/m - 1)$ is the net energy of a particle with magnetic moment M_0 in an applied field H , and the size summation starts at $m = 2$ to avoid the ill-defined interaction energy of an isolated particle. Although equation (1) resembles the grand canonical ensemble, in fact Γ contains only the intensive variables T , H and μ , allowing fluctuations in energy, magnetization and size. This is a type of 'generalized' ensemble that is rarely used in the macroscopic limit where the three intensive variables cannot be independent, but it is the only type of ensemble that does not artificially restrict the internal fluctuations of a bulk sample. Indeed, a central feature in the theory of nanothermodynamics^{6,7} is that small systems have large fluctuations so that only the appropriate ensemble exhibits correct behaviour.

The partition function, equation (1), yields all equilibrium thermal properties of the sample. Specifically, the average cluster size is $\bar{m} = \partial \ln \Gamma / \partial (\mu/k_B T)$, and the low-field susceptibility per particle $[(k_B T \partial^2 \ln \Gamma / \partial H^2) / \bar{m}]_{H \rightarrow 0}$ is:

$$\chi(T) = \frac{\chi_0/k_B T}{\bar{m} \Gamma} \sum_{m=2}^{\infty} \sum_{l=0}^m \frac{(2l-m)^2 m!}{(m-l)!l!} \exp\{-m[\epsilon_m(l) - \mu]/k_B T\} \quad (2)$$

Table 1 Thermodynamic parameters

Substance (references)	Published parameters			This work	
	T_C (K)	Θ (K)	γ	ϵ_0/k_B (K)	$\mu/k_B T + \ln 2$
EuO (23, 24)	69.59	81	1.29	75.0	-4.0×10^{-3}
Gd (25–27)	293.6	317	1.23	306.3	-1.3×10^{-3}
Co (28, 29)	1,388	1,415	1.21	1,511	-5.2×10^{-3}
Ni (30, 31)	627.2	654	1.29	647.8	-8.0×10^{-4}

Shown are Curie temperature (T_C), Weiss constant (Θ), critical exponent (γ), net interaction (ϵ_0/k_B) and effective chemical potential ($\mu/k_B T + \ln 2$) as determined from the magnetic susceptibility of europium oxide, gadolinium, cobalt and nickel.

Equation (2) has three adjustable parameters: χ_0 governs the amplitude of response, ϵ_0 gives the energy scale of interactions, while μ controls the entire distribution of clusters. In general $\mu/k_B T$ is weakly temperature dependent; for example, an ideal gas has $\mu/k_B T = \text{constant} - (3/2) \ln(T)$. Empirically, Fig. 1 shows that over a wide range of temperatures, the constant values of $\mu/k_B T$ given in Table 1 provide good agreement with the measured magnetic susceptibilities from four standard ferromagnets. This single parameter, which for $\mu/k_B T \approx -\ln(2)$ yields large but finite clusters, controls all of the deviations from the Curie–Weiss law around ϵ_0/k_B and all of the critical behaviour near T_C .

The phase transition can be interpreted using the scaled alignment variable, $L = 2l/m - 1$, so that in zero field the energy per particle becomes $\epsilon(L) = -(\epsilon_0/2)L^2$. Note that $|L|$ is the usual ferromagnetic order parameter, and that the energy is a maximum when there is no order $\epsilon(0) = 0$, and a minimum when the cluster is fully aligned $\epsilon(\pm 1) = -\epsilon_0/2$. Using Stirling's approximation for the factorials, and converting the sum over l to an integral over L , the partition function becomes $\Gamma \approx \sum_{m=2}^{\infty} \int_{-1}^{+1} \exp[-m\Omega(L)/k_B T] dL$, where the grand potential per particle is $\Omega(L) = \epsilon(L) - \mu + (1/2)k_B T\{(1+L)\ln[(1/2)(1+L)] + (1-L)\ln[(1/2)(1-L)]\}$. Figure 2 shows that $\Omega(L)/k_B T$ exhibits typical Landau-like behaviour, with a parabolic curve at $T \gg \epsilon_0/k_B$, that flattens about $L = 0$ at $T = \epsilon_0/k_B$, then develops two distinct minima at $T < \epsilon_0/k_B$. An infinite cluster would reside in the single minimum at $L = 0$ for $T > \epsilon_0/k_B$, then for $T < \epsilon_0/k_B$ its symmetry would break into one of the ordered minima. Finite clusters, however, have Boltzmann-weighted fluctuations over all values of L , so that the average local-order parameter $|\bar{L}| \approx \sum_{m=2}^{\infty} \int_{-1}^{+1} |L| \exp[-m\Omega(L)/k_B T] dL / \Gamma$ is non-zero at all temperatures. Similarly, even for $T \leq \epsilon_0/k_B$, this short-range order does not yield a net spontaneous magnetization in a macroscopic sample because the bulk behaviour is averaged over many clusters with all values of L . Near T_C where $\bar{m} \gg 100$, $|\bar{L}|$ increases rapidly with decreasing temperature so that it is difficult to distinguish this abrupt local ordering from the true ferromagnetic transition. At T_C , the solid curve in Fig. 2 shows that $\Omega(L)$ becomes zero for one specific value of $|L|$, where the divergent sum in Γ yields a macroscopic fraction of these clusters, much like the divergence for one specific value of momentum in Bose–Einstein condensation. It is this divergence that matches the measured critical behaviour of all four ferromagnetic materials shown in Fig. 1.

An infinite cluster has no other clusters with which to interact, so its equilibrium energy above T_C is $\lim_{m \rightarrow \infty} \epsilon_m(m/2) = 0$, as in the classical mean-field theory. Here I have shown that a macroscopic

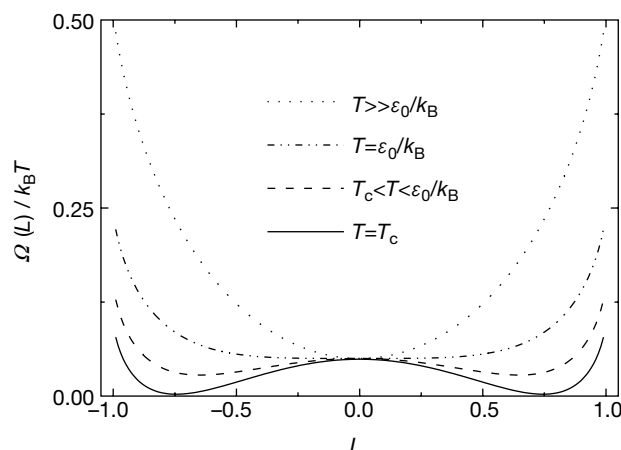


Figure 2 Landau-like plot of the reduced grand potential per particle, $\Omega(L)/k_B T$, as a function of the scaled alignment variable L , at four different temperatures. The offset from $\Omega(L)/k_B T = 0$ at $L = 0$, which keeps the partition function finite at $T > T_C$ and is shown in Table 1 to be $|\mu/k_B T + \ln 2| < 10^{-2}$, has been greatly amplified for clarity.

sample can increase its local order and lower its net energy by subdividing into finite clusters of strongly interacting particles, with relatively weak exchange of thermal properties between clusters. Furthermore, the sample can increase its entropy by forming aggregates of statistically indistinguishable clusters, with an average number of clusters per aggregate of $\bar{n} = \partial \ln(\sum_{n=1}^{\infty} \Gamma^n/n!)/\partial \ln \Gamma$ (ref. 5). Evidence for anomalous clustering can be found in measurements of magnetic hysteresis¹⁶, angle-resolved photoemission¹⁷ and neutron scattering¹⁸, but perhaps the most direct demonstration of thermodynamic heterogeneity comes from the technique of non-resonant spectral hole burning, NSHB¹⁹. Observation of NSHB requires the energy absorbed in a selected degree of freedom to remain localized for at least as long as its slow response, which is interpreted as selective local heating. Magnetic NSHB has been observed in a spin glass²⁰ and in single crystals of EuS and Fe (ref. 21).

The solid symbols in Fig. 1 show the measured magnetic correlation range (ξ) in crystalline Co (ref. 22). Assuming aggregates of compact clusters, a theoretical expression for ξ varies as $(\bar{m}\bar{n})^{1/3}$. Indeed, $(\bar{m}\bar{n})^{1/3}$ exhibits a temperature dependence that is similar to ξ , but its amplitude is about a factor of ten too large. At least part of this discrepancy comes from the inadequacy of the Ising model for real spins. When normalized to give one atomic radius as $T \rightarrow \infty$ (solid grey curve in Fig. 1), there is quantitative agreement between $(\bar{m}\bar{n})^{1/3}$ and ξ with no new parameters. Thus I have shown that the failure of classical mean-field theories to yield local correlations, and the failure of homogeneous models to yield the measured critical exponents, can be resolved by applying essentially the same mean-field theory to finite clusters with unrestricted sizes. □

Received 24 May; accepted 20 September 2000.

- Weiss, P. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. (Paris)* **6**, 661–690 (1907).
- Stanley, H. E. *Introduction to Phase Transitions and Critical Phenomena* (Oxford Univ. Press, New York, 1971).
- Stanley, H. E. Scaling universality, and renormalization. *Rev. Mod. Phys.* **71**, S358–S366 (1999).
- Collins, M. F. *Magnetic Critical Scattering* (Oxford Univ. Press, New York, 1989).
- Chamberlin, R. V. Mesoscopic mean-field theory for supercooled liquids and the glass transition. *Phys. Rev. Lett.* **82**, 2520–2523 (1999).
- Hill, T. L. *Thermodynamics of Small Systems* Parts I and II (Dover, New York, 1994).
- Hill, T. L. & Chamberlin, R. V. Extension of the thermodynamics of small systems to open metastable states: An example. *Proc. Natl Acad. Sci. USA* **95**, 12779–12782 (1998).
- Creutz, M. Deterministic Ising dynamics. *Ann. Phys.* **167**, 62–72 (1986).
- Rowlinson, J. S. Legacy of van der Waals. *Nature* **244**, 414–417 (1973).
- Onsager, L. Crystal statistics I. A two-dimensional model with an order-disorder transition. *Phys. Rev.* **65**, 117–149 (1944).
- Guggenheim, E. A. The principle of corresponding states. *J. Chem. Phys.* **13**, 253–261 (1945).
- Wegner, F. J. Corrections to scaling laws. *Phys. Rev. B* **5**, 4529–4536 (1972).
- Souletie, J. & Tholence, J. L. Critical behavior of nickel between T_C and $3T_C$. *Solid State Commun.* **48**, 407–410 (1983).
- Smart, J. S. *Effective Field Theories of Magnetism* (Saunders, Philadelphia, 1966).
- Huang, K. *Statistical mechanics* 2nd edn (Wiley, New York, 1987).
- Chamberlin, R. V. & Holtzberg, F. Remanent magnetization of a simple ferromagnet. *Phys. Rev. Lett.* **67**, 1606–1609 (1991).
- Korenman, V. Theories of itinerant magnetism. *J. Appl. Phys.* **57**, 3000–3005 (1985).
- Mezei, F., Farago, B., Hayden, S. M. & Stirling, W. G. Breakdown of conventional dynamic scaling at the ferromagnetic Curie point in EuO. *Physica B* **156&157**, 226–228 (1989).
- Schiener, B., Böhm, R., Loidl, A. & Chamberlin, R. V. Nonresonant spectral hole burning in the slow dielectric response of supercooled liquids. *Science* **274**, 752–754 (1996).
- Chamberlin, R. V. Nonresonant spectral hole burning in a spin glass. *Phys. Rev. Lett.* **83**, 5134–5137 (1999).
- Chamberlin, R. V. Experiments and theory of the nonexponential relaxation in liquids, glasses, polymers and crystals. *Phase Transitions* **65**, 169–209 (1998).
- Bally, D., Popović, M., Totia, M., Grabcev, B. & Lungu, A. M. Small-angle critical magnetic scattering of neutrons in Co. *Neutron Inelastic Scattering* Vol. II, 75–82 (IAEA, Vienna, 1968).
- Menyuk, N., Dwight, K. & Reed, T. B. Critical magnetic properties and exchange interactions in EuO. *Phys. Rev. B* **3**, 1689–1698 (1976).
- Huang, C. C. & Ho, J. T. Faraday rotation near the Curie point of EuO. *Phys. Rev. B* **12**, 5255–5260 (1975).
- Arajs, S. & Colvin, R. V. Paramagnetism of polycrystalline gadolinium, terbium, and dysprosium metals. *J. Appl. Phys.* **32**, 3365–3375 (1961).
- Nigh, H. E., Legvold, S. & Spedding, F. H. Magnetization and electrical resistivity of gadolinium single crystals. *Phys. Rev.* **132**, 1092–1097 (1963).
- Geldart, D. J. W., Hargreaves, P., Fujiki, N. M. & Dunlap, R. A. Anisotropy of the critical magnetic susceptibility of gadolinium. *Phys. Rev. Lett.* **62**, 2728–2731 (1989).
- Nakagawa, Y. Change of magnetic susceptibility of transition metals and alloys at their melting points. *J. Phys. Soc. Jpn.* **11**, 855–863 (1956).

- Colvin, R. V. & Arajs, S. Magnetic susceptibility of face-centred cubic cobalt just above the ferromagnetic Curie temperature. *J. Phys. Chem. Solids* **26**, 435–437 (1965).
- Fallot, M. Paramagnétisme des éléments ferromagnétiques. *J. Phys. (Paris)* **8**, 153–163 (1944).
- Arajs, S. Paramagnetic behavior of nickel just above the ferromagnetic Curie temperature. *J. Appl. Phys.* **36**, 1136–1137 (1965).

Acknowledgements

I thank K. E. Schmidt and G. H. Wolf for several insights into the subject reported here. I also thank B. Geil, T. L. Hill, S. M. Lindsay and R. Richert for discussions. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to the author (e-mail: ralph.chamberlin@asu.edu).

Universal quantum computation with the exchange interaction

D. P. DiVincenzo*, D. Bacon†‡, J. Kempe†§, G. Burkard¶ & K. B. Whaley†

* IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

† Department of Chemistry, ‡ Department of Physics, § Department of Mathematics, University of California, Berkeley, California 94720, USA

¶ École Nationale Supérieure des Télécommunications, 46 rue Barrault, 75634 Paris Cedex 13, France

¶ Department of Physics and Astronomy, University of Basel, Klingelbergstrasse 82, CH-4056 Basel, Switzerland

Various physical implementations of quantum computers are being investigated, although the requirements¹ that must be met to make such devices a reality in the laboratory at present involve capabilities well beyond the state of the art. Recent solid-state approaches have used quantum dots², donor-atom nuclear spins³ or electron spins⁴; in these architectures, the basic two-qubit quantum gate is generated by a tunable exchange interaction between spins (a Heisenberg interaction), whereas the one-qubit gates require control over a local magnetic field. Compared to the Heisenberg operation, the one-qubit operations are significantly slower, requiring substantially greater materials and device complexity—potentially contributing to a detrimental increase in the decoherence rate. Here we introduced an explicit scheme in which the Heisenberg interaction alone suffices to implement exactly any quantum computer circuit. This capability comes at a price of a factor of three in additional qubits, and about a factor of ten in additional two-qubit operations. Even at this cost, the ability to eliminate the complexity of one-qubit operations should accelerate progress towards solid-state implementations of quantum computation¹.

The Heisenberg interaction has many features^{2,5} that have led to its being chosen as the fundamental two-qubit interaction in a large number of recent proposals for the implementation of quantum computation. Its functional form is very accurate—deviations from the isotropic form of the interaction, arising only from relativistic corrections, can be very small in suitably chosen systems. It is a strong interaction, so that it should permit very fast gate operation, well into the GHz range for several of the proposals. At the same time, it is very short-ranged, arising from the spatial overlap of electronic wavefunctions, so that it should be possible to have an on–off ratio of many orders of magnitude. Unfortunately, the Heisenberg interaction by itself is not a universal gate⁶, in the sense that it cannot generate any arbitrary unitary transformation on a collection of spin-1/2 qubits. So, every proposal has supplemented the Heisenberg interaction with some other means of

applying independent one-qubit gates (which can be thought of as time-dependent local magnetic fields). But the need to add this capability to the device adds considerably to the complexity of the structures, by putting unprecedented demands on “g-factor” engineering of heterostructure materials^{4,7}—requiring that strong, inhomogeneous magnetic fields be applied^{2,5}—or involving microwave manipulations of the spins that may be slow and may cause heating of the device⁴. These added complexities may well exact a high cost, perhaps degrading the quantum coherence and clock rate of these devices by orders of magnitude.

The reason that the Heisenberg interaction alone does not give a universal quantum gate is that it has too much symmetry: it commutes with the operators $\hat{S}^2$ and $\hat{S}_z$ (for the total spin angular momentum and its projection on the z axis), and therefore it can only rotate among states with the same S , S_z quantum numbers. But by defining coded qubit states—ones for which the spin quantum numbers always remain the same—the Heisenberg interaction alone is universal^{8–10}, and single-spin operations with all their attendant difficulties can be avoided.

Recent work has identified the coding required to accomplish this. Early work identified techniques for suppressing phase-loss mechanisms due to coupling with the environment^{11–13}, and more recent studies have identified encodings that are completely immune from general collective decoherence—in which a single environmental degree of freedom couples in the same way to all the spins in a block. These codes are referred to both as decoherence-free subspaces (and their generalization, the decoherence-free subsystems)^{8,10,14}, and also as noiseless subspaces and subsystems^{9,15,16}. The noiseless properties of these codes are not relevant to the present work; but they have the desired property that they consist of states with definite angular-momentum quantum numbers.

So, in principle, the problem has been solved: the Heisenberg interaction alone is universal and can be used for quantum computation. However, a very practical question still remains: how great is the price that must be paid in return for eliminating single-spin operations? In particular, how many applications of the Heisenberg interaction are needed to complete some desired quan-

tum gate operation? The only guidance provided by the existing theory^{8–10} comes from a theorem of Solovay and Kitaev (refs 17, 18, and R. Solovay, personal communication), which states that “efficient” approximations exist: given a desired accuracy parameter ϵ , the number N of exchange operations required goes like $N \approx K \log(1/\epsilon)$, where $c \approx 4$ and K is an unknown positive constant. However, this theorem provides very little useful practical guidance for experiment; it does not show how to obtain the desired approximating sequence of exchange operations, and, as K is unknown, it gives no clue to whether the number of operations needed for a practical accuracy parameter is 10 or 10,000. Here we remedy these inadequacies by showing that the desired quantum logic operations can be obtained exactly, using sequences of exchange interactions short enough to be of practical significance for forthcoming experiments.

In the scheme we analyse here, we use the smallest subspace with definite angular-momentum quantum numbers that can be used to encode a qubit; this subspace is made up of three spins. We note¹⁰ that in principle the overhead in spatial resources could be made arbitrarily small: the rate of encoding into such noiseless subsystems converges asymptotically to unity. The space of three-spin states with spin quantum numbers $S = 1/2$, $S_z = +1/2$ is two-dimensional, and will serve to represent our coded qubit. A good explicit choice for the basis states of this qubit are $|0\rangle = |S\rangle|\uparrow\rangle$, $|1\rangle = (2/3)^{1/2}|T_+\rangle|\downarrow\rangle - (1/3)^{1/2}|T_0\rangle|\uparrow\rangle$. Here $|S\rangle = (1/2)^{1/2}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$ is the singlet state of spins 1 and 2 (see Fig. 1a) of the three-spin block, and $|T_+\rangle = |\uparrow\uparrow\rangle$ and $|T_0\rangle = (1/2)^{1/2}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$ are triplet states of these two spins. For these states we have constructed an explicit exchange implementation of the basic circuit elements of quantum logic⁶; in particular, we discuss how to obtain any coded one-qubit gate, and a specific two-qubit gate, the controlled NOT (CNOT).

We now show how one-qubit gates are performed on a single three-spin block. We write the Heisenberg hamiltonian as $H_{ij} = J \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j$ between spins i and j . Hamiltonian H_{12} generates a rotation $U_{12} = \exp(i\hbar^{-1} \int \mathbf{J} \mathbf{S}_1 \cdot \mathbf{S}_2 dt)$ which is just a z -axis rotation (in Bloch-sphere notation) on the coded qubit, while H_{23} produces a rotation about an axis in the x - z plane, at an angle of 120° from the z -axis. As simultaneous application of H_{12} and H_{23} can generate a rotation around the x -axis, three steps of one-dimensional (1D) parallel operation (defined in Fig. 1a) permit any one-qubit rotation, using the classic Euler-angle construction. In serial operation, we find numerically that four steps are always adequate when only nearest-neighbour interactions are possible (for example, the sequence $H_{23} - H_{12} - H_{23} - H_{12}$ shown in Fig. 2a, with suitable interaction strengths), while three steps suffice if interactions can be turned on between any pair of spins (for example, $H_{23} - H_{12} - H_{13}$; Fig. 2b).

We have performed numerical searches for the implementation of two-qubit gates using a simple minimization algorithm. Much of the difficulty of these searches arises from the fact that while the four basis states $|0_L, 1_L\rangle|0_L, 1_L\rangle$ have total spin quantum numbers $S = 1$, $S_z = +1$, the complete space with these quantum numbers for six spins has nine states, and exchanges involving these spins perform rotations in this full nine-dimensional space. So, for a given sequence, such as that depicted in Fig. 2c, we consider the resulting unitary evolution in this nine-dimensional Hilbert space as a function of the interaction times $t_1, t_2, \dots, t_N$. This unitary evolution can be expressed as a product $U(t_1, \dots, t_N) = U_N(t_N) \dots U_2(t_2) U_1(t_1)$, where $U_n(t_n) = \exp(it_n H_{i(n),j(n)}/\hbar)$. The objective of the algorithm is to find a set of interaction times such that the total time evolution describes a CNOT gate in the four-dimensional logic subspace $U(t_1, \dots, t_N) = U_{\text{CNOT}} \oplus A_5$. The matrix A_5 can be any unitary 5×5 matrix (consistent with U having a block diagonal form). The efficiency of our search is considerably improved by the use of two invariant functions

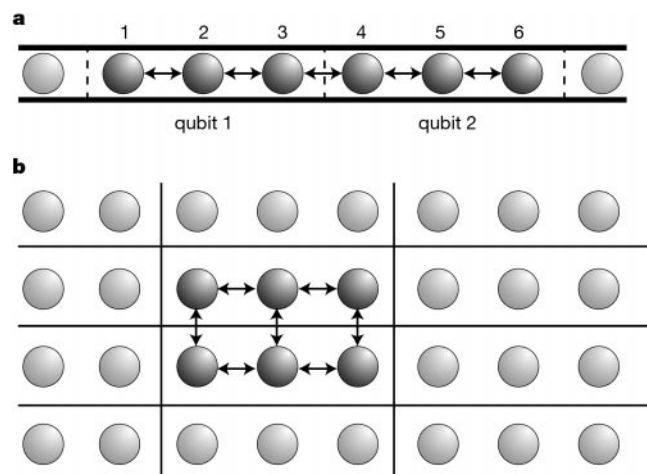


Figure 1 Possible layouts of spin-1/2 devices. **a**, One-dimensional layout. We consider two different assumptions about how the exchange interactions can be turned on and off in this layout: (1) at any given time each spin can be exchange-coupled to at most one other spin (we refer to this as “serial operation” in the text); (2) all exchange interactions can be turned on simultaneously between any neighbouring pair of spins in the line shown (“1D parallel operation”). **b**, Possible two-dimensional layout with interactions in a rectangular array. We imagine that any exchange interaction can be turned on between neighbouring spins in this array (“2D parallel operation”). Other arrangements are possible, but those shown here should be representative of the constraints that will be faced in actual device layouts.

$m_{1,2}(U)$ identified by Makhlin¹⁹, which are the same for any pair of two-qubit gates that are identical up to one-qubit rotations. It is then adequate to use an algorithm that searches for local minima of the function $f(t_1, \dots, t_N) = \sum_i (m_i(U_{\text{CNOT}}) - m_i(U(t_1, \dots, t_N)))^2$ with respect to $t_1, \dots, t_N$ (m_i is understood only to act on the 4×4 logic subspace of U). Finding a minimum for which $f = 0$ identifies an implementation of CNOT (up to additional one-qubit gates, which are easy to identify¹⁹) with the given sequence $(i(n), j(n))_n$, $i(n) \neq j(n)$ of exchange gates. If no minimum with $f = 0$ is found after many tries with different starting values (ideally mapping out all local minima), we have strong evidence (although not a mathematical proof) that the given sequence of exchange gates cannot generate CNOT.

The optimal serial-operation solution is shown in Fig. 2c. We note that this solution happens to involve only nearest neighbours in the 1D arrangement of Fig. 1a. The circuit layout shown has a high degree of symmetry; however, it does not appear possible to give the obtained solution in a closed form. (Any gate sequence involving non-nearest neighbours can be converted to a local gate sequence by swapping the involved qubits, using the SWAP gate, until they are close; but here the minimal solution found does not require such manipulations.) We have also found (apparently) optimal numerical solutions for parallel operation mode. For the 1D layout of Fig. 1a, the simplest solution found involves 8 clock cycles with just 8×4 different interaction-time parameters (H_{12} can always be zero in this implementation). For the 2D parallel mode of Fig. 1b, a solution was found using just 7 clock cycles (7×7

interaction times). The interaction parameters for these solutions are available from the authors on request.

It is worthwhile to give a complete overview of how quantum computation would proceed in the present scheme. It should begin by setting all the computational qubits to the $|0_L\rangle$ state. This state may be obtained using the exchange interaction: if a strong H_{12} is turned on in each coded block and the temperature made lower than the strength J of the interaction, these two spins will equilibrate to their ground state, which is the singlet state. The third spin in the block should be in the $|1\rangle$ state, which can be achieved by also placing the entire system in a moderately strong magnetic field B , such that $k_B T \ll g\mu_B B < J$. Then, computation can begin, with the one- and two-qubit gates implemented according to the schemes mentioned above. For the final qubit measurement, we note that determining whether the spins 1 and 2 of the block are in a singlet or a triplet suffices to perfectly distinguish $|0_L\rangle$ from $|1_L\rangle$ (again, the state of the third spin does not enter). Thus, for example, the a.c. capacitance scheme for spin measurement proposed by Kane³ is directly applicable to the coded-qubit measurement.

There are several issues raised by this work that deserve further exploration. The $S = 1/2$, $S_z = +1/2$ three-spin states that we use are a subspace of a decoherence-free subsystem that has been suggested for use in quantum computing by exchange interactions^{10,16}. Use of this full subsystem, in which the coded qubit can be in any mixture of the $S_z = +1/2$ and the corresponding $S_z = -1/2$ states, would offer immunity from certain kinds of interactions with the environment, and would not require any magnetic field to be present, even for initialization of the qubits.

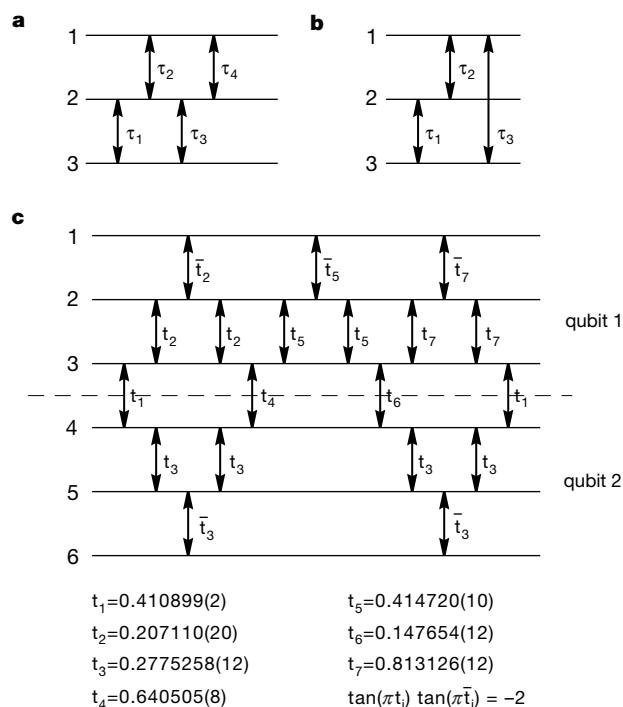


Figure 2 Circuits for implementing single-qubit and two-qubit rotations using serial operations. **a**, Single-qubit rotations by nearest-neighbour interactions. Four exchanges (double-headed arrows) with variable time parameters τ_i are always enough to perform any such rotation; one of the two possible layouts is shown. **b**, Non-nearest neighbour interactions. Only three interactions are needed; one of the possible layouts is shown. **c**, Circuit of 19 interactions that produce a CNOT between two coded qubits (up to one-qubit gates before and after). The durations of each interaction are given in units such that for $t = 1/2$ the rotation $U_{ij} = \exp(i/t \mathbf{S}_i \cdot \mathbf{S}_j / \hbar)$ is a SWAP, interchanging the quantum states of the two spins i, j . The $\bar{t}_i$ parameters are not independent, but are related to the t values as indicated. The uncertainty of the final digits of these times is indicated in parentheses. With these uncertainties, the absolute inaccuracy of the matrix elements of the two-qubit gate rotations achieved is no greater than 6×10^{-5} . Further fine-tuning of

these time parameters would give the CNOT to any desired accuracy. In a practical implementation, the exchange couplings $J(t)$ would be turned on and off smoothly; then the time values given here provide a specification for the integrated value $\int J(t) dt$. The functional form of $J(t)$ is irrelevant, but its integral must be controlled to the precision indicated. The numerical evidence is very strong that the solution shown here is essentially unique, so that no other choices of these times are possible, up to simple permutations and replacements $t \rightarrow 1 - t$ (we note that for the Heisenberg interaction adding any integer to t results in the same rotation). The results also strongly suggest that this solution is optimal: no one of these 19 interactions can be removed, and no other circuit layout with fewer than 19 has been found to give a solution. We have also sought, but not found, shorter implementation of other two-qubit gates like $\sqrt{\text{SWAP}}$ (refs 2, 5).

In this modified approach, the implementation of one-qubit gates is unchanged, but the CNOT implementation must satisfy additional constraints—the action of the exchanges on both the $S = 1$ and the $S = 0$ six-spin subspaces must be considered. As a consequence, implementation of CNOT in serial operation is considerably more complex; our numerical studies have failed to identify an implementation (even a good approximate one) for sequences of up to 36 exchanges (compare with 19 in Fig. 2c). On the other hand, we have found implementations using 8 clock cycles for 1D and 2D parallel operation (again for the 1D case H_{12} can be zero), so use of this larger Hilbert space may well be advantageous in some circumstances.

Finally, we note that further work is needed on the performance of quantum error correction within this scheme. Our logical qubits could be used directly within the error correction codes that have been shown to produce fault-tolerant quantum computation²⁰. Spin decoherence would primarily result in ‘leakage’ errors, which would take our logical qubits into states of different angular momentum (for example, $S = 3/2$). Our preliminary work indicates that, with small modifications, the conventional error correction circuits will not cause uncontrolled propagation of leakage error. In addition, the general theory^{8,10,20,21} shows that there exist sequences of exchange interactions that directly correct for leakage by swapping a fresh $|0_L\rangle$ into the coded qubit if leakage has occurred, and doing nothing otherwise; we have not yet identified numerically such a sequence. If fast measurements are possible, teleportation schemes could also be used in leakage correction.

The present results offer an alternative route to the implementation of quantum computation. The trade-offs are clear: for the price of a factor of three more devices, and about a factor of ten more clock cycles, the need for stringent control of magnetic fields applied to individual spins is dispensed with. We hope that the flexibility offered by these results will make easier the hard path to the implementation of quantum computation in the laboratory. □

Received 30 May; accepted 2 October 2000.

- DiVincenzo, D. P. The physical implementation of quantum computation. *Fortsch. Phys.* (special issue: Experimental Proposals for Quantum Computation) (in the press); also as preprint quant-ph/0002077 at (<http://xxx.lanl.gov>) (2000).
- Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
- Kane, B. E. A silicon-based nuclear-spin quantum computer. *Nature* **393**, 133–137 (1998).
- Vrijen, R. *et al.* Electron-spin-resonance transistors for quantum computing in silicon-germanium heterostructures. *Phys. Rev. A* **62**, 012306-1 to 012306-10 (2000).
- Burkard, G., Loss, D. & DiVincenzo, D. P. Coupled quantum dots as quantum gates. *Phys. Rev. B* **59**, 2070–2078 (1999).
- Barenco, A. *et al.* Elementary gates for quantum computation. *Phys. Rev. A* **52**, 3457–3467 (1995).
- DiVincenzo, D. P. *et al.* Quantum computation and spin electronics. In *Quantum Mesoscopic Phenomena and Mesoscopic Devices in Microelectronics* (eds Kulik, I. O. & Ellialtıoglu, R.) (NATO ASI, in the press), also as preprint cond-mat/9911245 at (<http://xxx.lanl.gov>) (2000).
- Bacon, D., Kempe, J., Lidar, D. A. & Whaley, K. B. Universal fault-tolerant computation on decoherence-free subspaces. *Phys. Rev. Lett.* **85**, 1758–1761 (2000).
- Viola, L., Knill, E. & Lloyd, S. Dynamical generation of noiseless quantum subsystems. *Phys. Rev. Lett.* **85**, 3520–3523 (2000).
- Kempe, J., Bacon, D., Lidar, D. A. & Whaley, K. B. Theory of decoherence-free fault-tolerant universal quantum computation. *Phys. Rev. A* (submitted); also as preprint quant-ph/0004064 at (<http://xxx.lanl.gov>) (2000).
- Zurek, W. H. Environment-induced superselection rules. *Phys. Rev. D* **26**, 1862–1880 (1982).
- Palma, G. M., Suominen, K.-A. & Ekert, A. K. Quantum computers and dissipation. *Proc. R. Soc. Lond. A* **452**, 567–584 (1996).
- Duan, L.-M. & Guo, G.-C. Reducing decoherence in quantum-computer memory with all quantum bits coupling to the same environment. *Phys. Rev. A* **57**, 737–741 (1998).
- Lidar, D. A., Chuang, I. L. & Whaley, K. B. Decoherence-free subspaces for quantum computation. *Phys. Rev. Lett.* **81**, 2594–2597 (1998).
- Zanardi, P. & Rasetti, M. Error avoiding quantum codes. *Mod. Phys. Lett. B* **11**, 1085–1093 (1997).
- Knill, E. & Laflamme, R. & Viola, L. Theory of quantum error correction for general noise. *Phys. Rev. Lett.* **84**, 2525–2528 (2000).
- Kitaev, A. Y. Quantum computations: algorithms and error correction. *Russ. Math. Surv.* **52**, 1191–1249 (1997).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* Appendix 3 (Cambridge Univ. Press, Cambridge, 2000).
- Makhlin, Y. Nonlocal properties of two-qubits gates and mixed states and optimization of quantum computations. Preprint quant-ph/0002045 at (<http://xxx.lanl.gov>) (2000).
- Preskill, J. in *Introduction to Quantum Computation and Information* (eds Lo, H.-K., Popescu, S. & Spiller, T.) 213–269 (World Scientific, Singapore, 1998).

21. Lidar, D. A., Bacon, D. & Whaley, K. B. Concatenating decoherence-free subspaces with quantum error correcting codes. *Phys. Rev. Lett.* **82**, 4556–4559 (1999).

Acknowledgements

We thank P. O. Boykin and B. M. Terhal for discussions. D.P.D., D.B., J.K. and K.B.W. were supported by the National Security Agency and the Advanced Research and Development Activity. D.P.D. also thanks the UCLA DARPA program on spin-resonance transistors for support, and is also grateful for the hospitality of D. Loss at the University of Basel, where much of this work was completed. J.K. also acknowledges support from the US National Science Foundation. G.B. is supported in part by the Swiss National Science Foundation.

Correspondence and requests for materials should be addressed to D.P.D. (e-mail: divince@watson.ibm.com).

Kondo physics in carbon nanotubes

Jesper Nygård*, David Henry Cobden† & Poul Erik Lindelof*

* Ørsted Laboratory, Niels Bohr Institute, Universitetsparken 5, DK-2100 Copenhagen, Denmark

† Department of Physics, University of Warwick, Coventry CV4 7AL, UK

The connection of electrical leads to wire-like molecules is a logical step in the development of molecular electronics, but also allows studies of fundamental physics. For example, metallic carbon nanotubes¹ are quantum wires that have been found to act as one-dimensional quantum dots^{2,3}, Luttinger liquids^{4,5}, proximity-induced superconductors^{6,7} and ballistic⁸ and diffusive⁹ one-dimensional metals. Here we report that electrically contacted single-walled carbon nanotubes can serve as powerful probes of Kondo physics, demonstrating the universality of the Kondo effect. Arising in the prototypical case from the interaction between a localized impurity magnetic moment and delocalized electrons in a metallic host, the Kondo effect has been used to explain¹⁰ enhanced low-temperature scattering from magnetic impurities in metals, and also occurs in transport through semiconductor quantum dots^{11–18}. The far greater tunability of dots (in our case, nanotubes) compared with atomic impurities renders new classes of Kondo-like effects^{19,20} accessible. Our nanotube devices differ from previous systems in which Kondo effects have been observed, in that they are one-dimensional quantum dots with three-dimensional metal (gold) reservoirs. This allows us to observe Kondo resonances for very large electron numbers (N) in the dot, and approaching the unitary limit (where the transmission reaches its maximum possible value). Moreover, we detect a previously unobserved Kondo effect, occurring for even values of N in a magnetic field.

Each of our devices²¹ contains a metallic single-walled nanotube with gold source and drain contacts and a substrate gate contact, as shown in Fig. 1a. Several measures are taken to optimize electrical contact between the gold and the nanotubes (see Methods). The resulting two-terminal linear-response conductance G ranges up to more than $3e^2/h$ (resistance $8\text{ k}\Omega$) at room temperature, where h is Planck’s constant and e is the electronic charge. We believe this is the highest reported yet for a single-walled tube²². This implies that the transmission probability P of each contact can reach about 0.9 (the maximum theoretical conductance of a metallic tube is $4e^2/h$), demonstrating that nearly ideal contacts are achievable between a normal metal and a molecule. Meanwhile, the variation of P from device to device allows the investigation of different transport regimes.

For $P \ll 1$, as is well established^{2,3}, Coulomb blockade occurs at low temperature T and the device behaves as a one-dimensional (1D) quantum dot²³. The values of the charging energy, U (the average interaction between two electrons), and the average level spacing, ΔE , found for this type of device are as expected for a tube

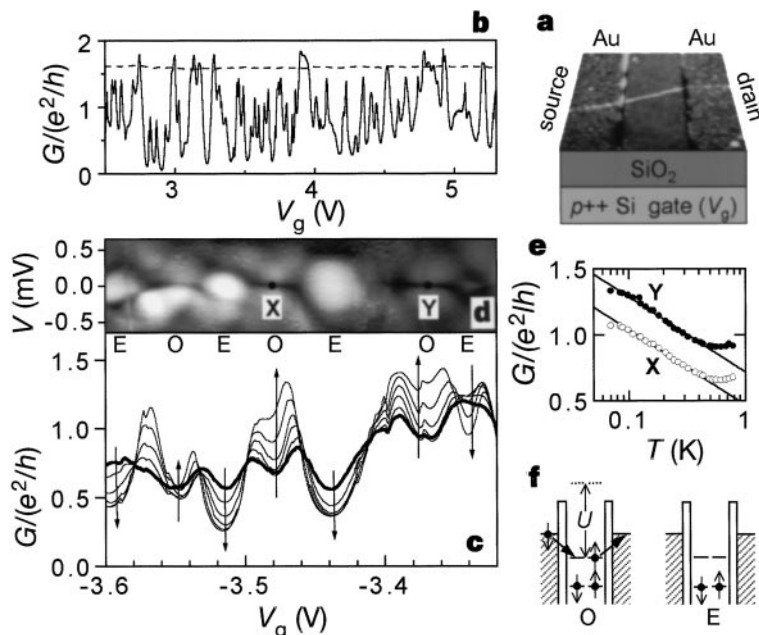


Figure 1 Characteristics of a nanotube device with intermediate contact transmission probabilities. **a**, Device consisting of a 2-nm-thick nanotube bundle with 30-nm-thick gold contacts separated by 300 nm. Sketched beneath the tapping-mode atomic force microscope (AFM) image are the 300-nm-thick SiO₂ layer and the highly *p*-doped Si substrate used as the gate electrode. **b**, At room temperature (dashed line) the linear-response conductance *G* is about 1.6 *e*²/*h*, almost independent of gate voltage *V_g*. At *T* ≈ 75 mK, the base electron temperature on our dilution refrigerator, reproducible fluctuations are seen in *G* versus *V_g*. (The source-drain bias *V* used was 7 μV d.c.).

c, Temperature dependence of a narrower region of *V_g*. *T* = 75, 125, 180, 245, 320, 490, 560 and (thicker line) 780 mK. Vertical arrows indicate the direction of change with decreasing *T*, which is opposite in the regions labelled E and O. **d**, Greyscale plot of *dI/dV* against *V* and *V_g*, over the same region of *V_g* (darker is more positive). **e**, Conductance versus temperature at points X and Y in **d**. The straight lines indicate log(*T*) behaviour. **f**, Interpretation of the situation in the O and E regions. In O regions the number of electrons *N* on the dot is odd, and higher-order spin-flipping processes can occur, leading to a Kondo resonance. In the E regions *N* is even and no such processes exist.

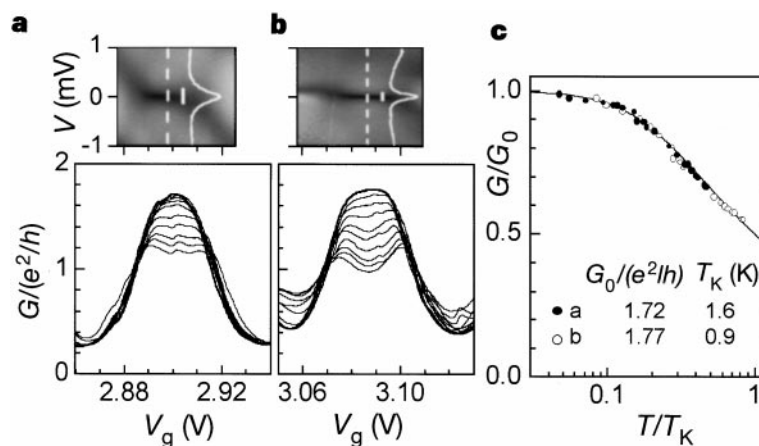


Figure 2 Analysis of two high-conductance peak pairs. **a**, **b**, Temperature dependence of *G* versus *V_g*. *T* = 75 (thick line), 87, 150, 200, 250, 330, 420, 530, 630 and 740 mK. Above are *dI/dV* greyscales at 75 mK in the same regions of *V_g*, showing broad Kondo ridges. Superimposed on each is the *dI/dV* versus *V* trace along the dashed line, which

also serves to denote *dI/dV* = 0. The vertical white bars are of length 2*k_BT_K*/*e*, where *T_K* is obtained in each case from the data in **c**. **c**, *G* versus *T* data in the valley centres for the peak pairs in **a** and **b**, scaled in each case to match the theoretical scaling function given in the text (solid line). The deduced values of *G₀* and *T_K* are given.

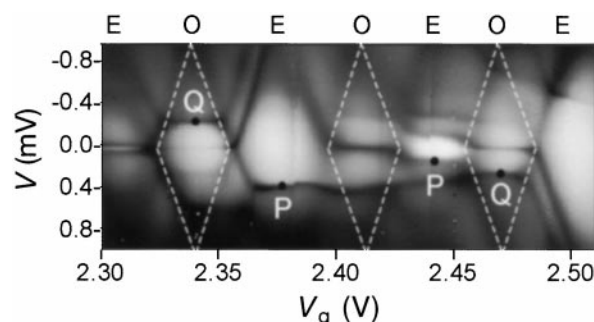


Figure 3 *dI/dV* greyscale plot in a different *V_g* region at *T* = 75 mK. Dashed lines outline the odd Coulomb diamonds. Horizontal features such as those labelled P and Q suggest

higher-order processes like the second-order one sketched on the right, where tunnelling involves two levels separated by an energy Δ and occurs only for bias $V \geq \Delta/e$.

segment of effective length L equal to the contact spacing³. At higher T , the device exhibits a power-law variation of G with T , which is well explained by Luttinger-liquid behaviour⁴. At the other extreme, for $P \approx 0.9$, we see no Coulomb blockade, and instead the characteristics²¹ resemble those of a diffusive 1D wire even for $T < 1$ K.

It is the intermediate transmission regime that we focus on here. The measurements in Fig. 1 were made on a device having $G \approx 1.6 e^2/h$ (we estimate $P \approx 0.6$) at room temperature. At low T , G undergoes large, reproducible fluctuations as a function of substrate gate voltage V_g (Fig. 1b). Figure 1c shows the variation with T over a narrower range of V_g . At $T = 780$ mK, the fluctuations are regular and can be identified as Coulomb blockade oscillations. However, unlike normal Coulomb blockade, as T is decreased G decreases in some valleys, marked above with an E, while it increases in other valleys, marked with an O. Figure 1d shows a greyscale plot of the differential conductance dI/dV versus source-drain bias V and V_g at $T \approx 75$ mK. The enhanced conductance in the O valleys corresponds to the dark (high dI/dV) horizontal features at $V = 0$, such as those labelled X and Y. No such lines are seen in the E valleys, which are instead bounded by roughly hexagonal bubbles. Although the distinctness of these features varies with V_g , O-type and E-type valleys can be seen to alternate for as many as ten consecutive peaks.

These features can be explained by invoking the Kondo effect^{11–13}. When the number N of electrons on the dot is odd, the total spin S is a half-integer, and second-order processes such as that shown on the left of Fig. 1f can occur. Similar processes, which change the total

spin of the dot, add up coherently to form a correlated many-electron state in which electrons in the two leads are strongly coupled, allowing current flow even under blockade conditions if $T < T_K$, the Kondo temperature. When N is even, however, there is no equivalent process, and thus no current, if $S = 0$. This matches the data if in the O valleys N is odd and $S = \frac{1}{2}$ while in the E valleys N is even and $S = 0$. Further evidence confirming that features X and Y in Fig. 1d are “Kondo ridges” is provided by the logarithmic dependence¹⁶ of G on T in the centres of the ridges, shown in Fig. 1e, and by the splitting of the ridges in a magnetic field (see later).

This new Kondo system differs in several important ways from the two dimensional (2D) semiconductor dots in which Kondo physics was previously studied^{14–19}. First, for semiconductor dots the leads are 2D electron systems, while for tube dots they are normal three-dimensional (3D) metal (gold). Second, the excitation spectrum of an interacting 1D dot, which behaves as a Luttinger liquid when Coulomb blockade is overcome⁴, is profoundly different from that of a 2D dot. Third, because semiconductor dots are electrostatically defined, their geometry and contact transmission probabilities depend on gate voltages. In contrast, for tube dots the contacts cannot be individually tuned, but the fixed geometry means that signs of the Kondo effect can be seen over the entire range of V_g , encompassing hundreds of Coulomb oscillations. Fourth, the effect of a magnetic field is entirely through spin in single-walled nanotube dots, whereas orbital effects are often dominant in semiconductor dots. Fifth, because of the latter, combined with the absence of spin-orbit coupling, unlike the spin of a semiconductor dot, the spin of a tube dot is well defined and

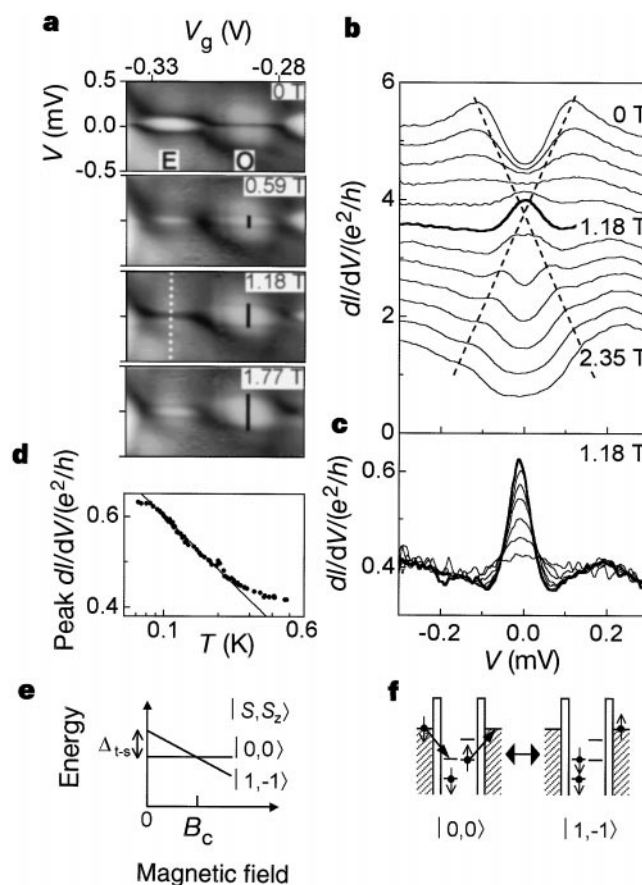


Figure 4 Effect of perpendicular magnetic field B . **a**, dI/dV greyscale plot at $T = 75$ mK showing adjacent odd- N (O) and even- N (E) regions at a series of fields. The vertical bars have length $2g\mu_B B/e$. As B increases the O ridge splits while the E bubble shrinks to a single ridge at 1.18 T before reappearing at higher B . **b**, Evolution of dI/dV versus V characteristics at $V_g = -0.322$ V (indicated by the white dotted line in **a**) at $T = 75$ mK. The traces are offset from each other by $0.4 e^2/h$ for clarity. The dashed

lines, which indicate the motion of the peaks, are sloped according to $g\mu_B B/e$. **c**, Temperature dependence of the peak in dI/dV at $B = 1.18$ T. $T = 75$ (thick line), 100, 115, 130, 180, 230 and 350 mK. **d**, Peak value of dI/dV (at $V = 0$) plotted against $\log(T)$. **e**, A singlet $|S, S_z\rangle = |0, 0\rangle$ ground state becomes degenerate with a triplet $|1, -1\rangle$ excited state at $B = B_c = \Delta_{t-s}/g\mu_B$. At this point, spin-flipping higher order transitions, as sketched in **f**, become possible, leading to a new type of Kondo resonance.

easily measured^{24,25} even for large N . Sixth, the conditions^{14,23} on the size L for observing the Kondo effect are weaker in the tube dot. Halfway along a Kondo ridge, $T_K \approx (\Gamma U)^{1/2} \exp[-(\pi/4)U/\Gamma]$, where Γ is the level width. The ratio U/Γ cannot be made smaller than $U/\Delta E$, because $\Gamma < \Delta E$ for any quantum dot. For 2D semiconductor dots, $U/\Delta E \propto L$, so that the maximum possible T_K decreases exponentially with L . This requires them to be made as small as possible, which limits the electron number N to less than about 100. In contrast, for tube dots, which are 1D, $U/\Delta E \approx 6$ is independent³ of length L , and the maximum T_K decreases only with the prefactor $(\Gamma U)^{1/2}$. This helps to explain how tube dots can show such a strong and clear Kondo effect even for large N . An order of magnitude estimate of N is the number of π -electrons in the tube, that is, the number of carbon atoms, which is tens of thousands. (Without knowing the exact dot size or band structure a more accurate estimate is not possible.) Nanotube quantum dots therefore allow the first ever observations, to our knowledge, of the Kondo effect in the limit of very large N .

The width Γ varies from level to level, so we can search for pairs of peaks with particularly large Γ , such as those shown in Fig. 2a and b. For both of these pairs the valley between the peaks is completely gone at base temperature (~ 75 mK). In the Kondo regime, G is expected to saturate at the unitary limit G_0 as $T \rightarrow 0$, and at higher T it should follow a universal scaling form¹⁰ that can be approximated by¹⁵ $G(T) = G_0/(1 + (2^{1/s} - 1)(T/T_K)^s)^2$, where $s = 0.22$ for spin half on the dot. To test this, we fit G versus T in the valley centre to this expression, using T_K and G_0 as fitting parameters. The results, shown in Fig. 2c, are highly satisfactory. For both peak pairs we find $G_0 > 1.7e^2/h$, which is close to the maximum value of $2e^2/h$ obtained for symmetric coupling to the two contacts (this assumes the double Fermi-point degeneracy¹ is broken). The deduced values of T_K of 1.6 K and 0.9 K are in good agreement with nonlinear measurements, as indicated in the insets above Fig. 2a and b. Here, on top of greyscale plots of dI/dV showing the Kondo ridges are superimposed traces of dI/dV versus V midway along each ridge. The Kondo effect is expected to be suppressed¹³ on a bias scale $|V| \approx k_B T_K/e$. Indeed, we see that the width of each peak in dI/dV is roughly $2k_B T_K/e$ (indicated by the vertical white bars). We believe these measurements on a nanotube dot are closer to the unitary limit than has previously been reported in a semiconductor quantum dot.

In some regions of V_g , such as that shown in Fig. 3, a regular series of faint diamonds can be discerned. This resembles the standard behaviour of a quantum dot for $P \ll 1$, where only first-order tunnelling is significant and Coulomb blockade forces N to be fixed at an integer within the diamonds. However, superimposed on the diamonds are a variety of horizontal features which can be attributed to higher-order tunnelling processes that are strong in this device because P is large. Kondo ridges are seen at $V = 0$ in the odd- N diamonds (marked O again), which are outlined by white dashed lines. Additional horizontal ridges are seen at $V \neq 0$, such as those labelled P in the even and Q in the odd diamonds. Features of type P truncate every even diamond, resulting in the characteristic alternating bubble-ridge pattern seen also in Fig. 1d. It is natural to associate a horizontal feature at a bias $V = \Delta/e$ with processes generating an excitation of energy Δ in the dot. An example of such a process is co-tunnelling involving two nondegenerate single-particle states, as sketched on the right of Fig. 3. However, since such inelastic co-tunnelling normally only produces a step in dI/dV above a threshold, we suggest that these peaks in dI/dV at finite bias offsets are another manifestation of Kondo physics.

Finally, we look at the effect of magnetic field B . Figure 4a shows the evolution with B of an adjacent bubble (E) and Kondo ridge (O). As expected¹³, the Kondo ridge splits linearly into components at $V = \pm g\mu_B B/e$ (the Zeeman energy), where μ_B is the Bohr magneton and $g = 2.0$ is the electron g -factor^{2,25}. Meanwhile, the edges of the bubble move linearly towards $V = 0$. Figure 4b shows the dI/dV

characteristics in the centre of the bubble. At a certain field, $B_c = 1.18$ T, the bubble collapses to a single ridge, corresponding to a peak in dI/dV at $V = 0$. Its T dependence is shown in Fig. 4c. The peak height is logarithmic in T (Fig. 4d), just as for the odd- N Kondo ridge at $B = 0$. A similar ridge is formed in most bubbles at a field dependent on the bubble width at $B = 0$.

Exactly such behaviour has recently been predicted²⁰ for quantum dots when N is even, in the case where the Zeeman energy is sufficiently large. It is assumed that at $B = 0$ the ground state is a singlet ($S = 0$), and the lowest excited state is a triplet ($S = 1$) at a distance Δ_{t-s} above the ground state. In an applied field $B_c = \Delta_{t-s}/g\mu_B$, the $S_z = -1$ member of the triplet becomes degenerate with the singlet (see Fig. 4e). At this point, S alternates between 1 and 0 as electrons co-tunnel between the contacts (see Fig. 4f), resulting in a new type of Kondo resonance²⁰ with similar characteristics to the conventional type seen at $B = 0$. This theory is particularly appropriate to single-walled nanotubes, where the effects of B are almost entirely through spin. Applying the model, we deduce for instance that the edges of the E bubble in Fig. 4a are associated with a singlet-triplet excitation of energy $\Delta_{t-s} = g\mu_B B_c = 137 \mu\text{eV}$.

We note that this new type of Kondo effect is distinct from the one recently discovered in a semiconductor dot¹⁹, which occurs when a singlet is aligned with a degenerate triplet by orbital effects in a weak magnetic field. It serves to underline the fact that new physical phenomena can emerge in the study of molecular nanostructures. □

Methods

In making devices, the as-grown single-walled nanotube material²⁶ is sonicated in dichloroethane and deposited on the SiO_2 by diluting with isopropyl alcohol before drying. To improve the contacts: the leads are fabricated as soon as possible after tube deposition by low-energy (10 kV) electron-beam lithography; the polymethylmethacrylate resist is over-exposed to avoid residue; the contact metallization is pure gold, evaporated directly on top of the tubes; and each contact covers at least $0.5 \mu\text{m}$ of the tube or bundle. Although we cannot resolve the difference between a single tube and a thin bundle of tubes in the atomic force microscope, we study only devices whose detailed characteristics imply that the conduction is determined by a single metallic nanotube²¹.

Received 30 May; accepted 3 October 2000.

- Dekker, C. Carbon nanotubes as molecular quantum wires. *Phys. Today* **52**, 22–28 (1999).
- Tans, S. *et al.* Individual single-walled carbon nanotubes as quantum wires. *Nature* **386**, 474–477 (1997).
- Bockrath, M. *et al.* Single-electron transport in ropes of carbon nanotubes. *Science* **275**, 1922–1925 (1997).
- Bockrath, M. *et al.* Luttinger-liquid behaviour in carbon nanotubes. *Nature* **397**, 598–601 (1999).
- Yao, Z., Postma, H. W. C., Balents, L. & Dekker, C. Carbon nanotube intramolecular junctions. *Nature* **402**, 273–276 (1999).
- Kasumov, A. Y. *et al.* Supercurrents through single-walled carbon nanotubes. *Science* **284**, 1508–1511 (1999).
- Morpurgo, A. F., Kong, J., Marcus, C. & Dai, H. Gate-controlled superconducting proximity effect in carbon nanotubes. *Science* **286**, 263–265 (1999).
- Frank, S., Poncharal, S. P., Wang, Z. L. & de Heer, W. A. Carbon nanotube quantum resistors. *Science* **280**, 1744–1746 (1998).
- Bachtold, A. *et al.* Aharonov–Bohm oscillations in carbon nanotubes. *Nature* **397**, 673–675 (1999).
- Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).
- Glazman, I. I. & Raikh, M. E. Resonant Kondo transparency of a barrier with quasilocal impurity states. *JETP Lett.* **47**, 452–455 (1988).
- Ng, T. K. & Lee, P. A. On-site Coulomb repulsion and resonant tunneling. *Phys. Rev. Lett.* **61**, 1768–1771 (1988).
- Meir, Y., Wingreen, N. S. & Lee, P. A. Low-temperature transport through a quantum dot: the Anderson model out of equilibrium. *Phys. Rev. Lett.* **70**, 2601–2604 (1993).
- Goldhaber-Gordon, D. *et al.* Kondo effect in a single-electron transistor. *Nature* **391**, 156–159 (1998).
- Goldhaber-Gordon, D. *et al.* From the Kondo regime to the mixed-valence regime in a single-electron transistor. *Phys. Rev. Lett.* **81**, 5225–5228 (1998).
- Cronenwett, S. M., Oosterkamp, T. H. & Kouwenhoven, L. P. A tuneable Kondo effect in quantum dots. *Science* **281**, 540–544 (1998).
- Schmid, J., Weis, J., Eberl, K. & Klitzing, K. A quantum dot in the limit of strong coupling to reservoirs. *Physica B* **256–258**, 182–185 (1998).
- Simmel, F., Blick, R. H., Kotthaus, J. P., Wegscheider, W. & Bichler, M. Anomalous Kondo effect in a quantum dot at nonzero bias. *Phys. Rev. Lett.* **83**, 804–807 (1999).
- Sasaki, S. *et al.* A novel Kondo effect in an integer-spin quantum dot. *Nature* **405**, 764–767 (2000).
- Pustilnik, M., Avishai, Y. & Kikoin, K. Quantum dots with even number of electrons: Kondo effect in a finite magnetic field. *Phys. Rev. Lett.* **84**, 1756–1759 (2000).
- Nygård, J., Cobden, D. H., Bockrath, M., McEuen, P. L. & Lindelof, P. E. Electrical transport measurements on single-walled carbon nanotubes. *Appl. Phys. A* **69**, 297–304 (1999).
- Soh, H. T. *et al.* Integrated nanotube circuits: Controlled growth and ohmic contacting of single-walled carbon nanotubes. *Appl. Phys. Lett.* **75**, 627–629 (1999).

23. Glazman, L. I. Single electron tunneling. *J. Low Temp. Phys.* **118**, 247–269 (2000).
24. Tans, S., Devoret, M. H., Groeneveld, R. J. A. & Dekker, C. Electron–electron correlations in carbon nanotubes. *Nature* **394**, 761–764 (1998).
25. Cobden, D. H., Bockrath, M., McEuen, P. L., Rinzler, A. G. & Smalley, R. E. Spin splitting and even-odd effects in carbon nanotubes. *Phys. Rev. Lett.* **81**, 681–684 (1998).
26. Thess, A. *et al.* Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–487 (1996).

Acknowledgements

We thank A. Rinzler and R. Smalley for supplying the nanotubes, K. G. Rasmussen, M. M. Andreasen, A. E. Hansen and A. Kristensen for experimental assistance, and M. Pustilnik, N. Wingreen, L. P. Kouwenhoven, N. d'Ambrumenil, P. R. Poulsen and P. L. McEuen for helpful discussions.

Correspondence should be addressed to D.H.C. (e-mail: d.h.cobden@warwick.ac.uk).

Increased marine production of N_2O due to intensifying anoxia on the Indian continental shelf

S. W. A. Naqvi, D. A. Jayakumar*, P. V. Narvekar, H. Naik, V. V. S. S. Sarma*, W. D'Souza, S. Joseph & M. D. George

National Institute of Oceanography, Dona Paula, Goa 403 004, India

Eutrophication of surface waters and hypoxia in bottom waters has been increasing in many coastal areas^{1–4}, leading to very large depletions of marine life in the affected regions⁴. These areas of high surface productivity and low bottom-water oxygen concentration are caused by increasing runoff of nutrients from land. Although the local ecological and socio-economic effects have received much attention^{2–4}, the potential contribution of increasing hypoxia to global-change phenomena is unknown. Here we report the intensification of one of the largest low-oxygen zones in the ocean, which develops naturally over the western Indian continental shelf during late summer and autumn. We also report the highest accumulations yet observed of hydrogen sulphide (H_2S) and nitrous oxide (N_2O) in open coastal waters. Increased N_2O production is probably caused by the addition of anthropogenic nitrate and its subsequent denitrification, which is favoured by hypoxic conditions. We suggest that a global expansion of hypoxic zones may lead to an increase in marine production and emission of N_2O , which, as a potent greenhouse gas, could contribute significantly to the accumulation of radiatively active trace gases in the atmosphere⁵.

The coastal zone off western India experiences moderate upwelling during June–November^{6,7}. However, the cold, saline upwelled water is usually capped by a 5–10-m thick warm, lower-salinity layer arising from large land runoff and local precipitation, causing strong stratification and poor ventilation of sub-pycnocline waters. The upwelled water is derived from a poleward undercurrent, located just off the shelf break⁸, which has an oxygen (O_2) content of about 0.5 ml l^{-1} ($22\text{ }\mu\text{M}$) at 15°N . Over the shelf, the O_2 content is quickly depleted to near-zero levels owing to excessive consumption fuelled by high primary production mainly within the pycnocline ($>500\text{ mg C m}^{-3}\text{ d}^{-1}$)⁹. During September–October 1999, severely hypoxic ($O_2 < 22\text{ }\mu\text{M}$) waters were found over almost the entire shelf (Fig. 1), covering an area of about $180,000\text{ km}^2$ —an order of magnitude greater than the area

($20,000\text{ km}^2$) of the largest human-induced hypoxic zone in the Gulf of Mexico³.

Sub-pycnocline O_2 depletion over the western Indian shelf, as in other similar environments off Namibia¹⁰ and Peru¹¹, is primarily of natural origin because the nutrient enrichment occurs mainly through upwelling; however, it has not been known to be as severe as observed in the present study. For example, observations in October 1999 along a cross-shelf transect off Bombay revealed the prevalence of reducing conditions over much of the shelf with a progression from hypoxic (nonreducing; NO_2^- [nitrite] = 0, $O_2 > 4.4\text{ }\mu\text{M}$) to suboxic (denitrifying; NO_3^- , $\text{NO}_2^- > 0$, $0 < O_2 < 4.4\text{ }\mu\text{M}$) and then to anoxic (sulphate reducing; NO_3^- , NO_2^- , $O_2 = 0$) regimes as the upwelled water moved up from the shelf-break to the inner-shelf (Fig. 2). These conditions start developing in June, reach peak intensity by September–October and dissipate by December. Thus, this shallow, seasonal suboxic zone is distinct from the deeper, perennial suboxic layer of the central Arabian Sea¹².

On three of our five cruises during the upwelling period, H_2S was frequently detected in concentrations reaching up to $19\text{ }\mu\text{M}$ over the inner-shelf north of about 12°N latitude as well as on numerous occasions along the shorter transect off Goa. Instances of sulphate reduction in open-oceanic waters have been recorded off Peru (often without H_2S analysis)^{11,13}, but never before off India. This is presumably because although the O_2 levels were low, the water was not completely anoxic. The earliest data sets (in the 1950s, refs 6, 7

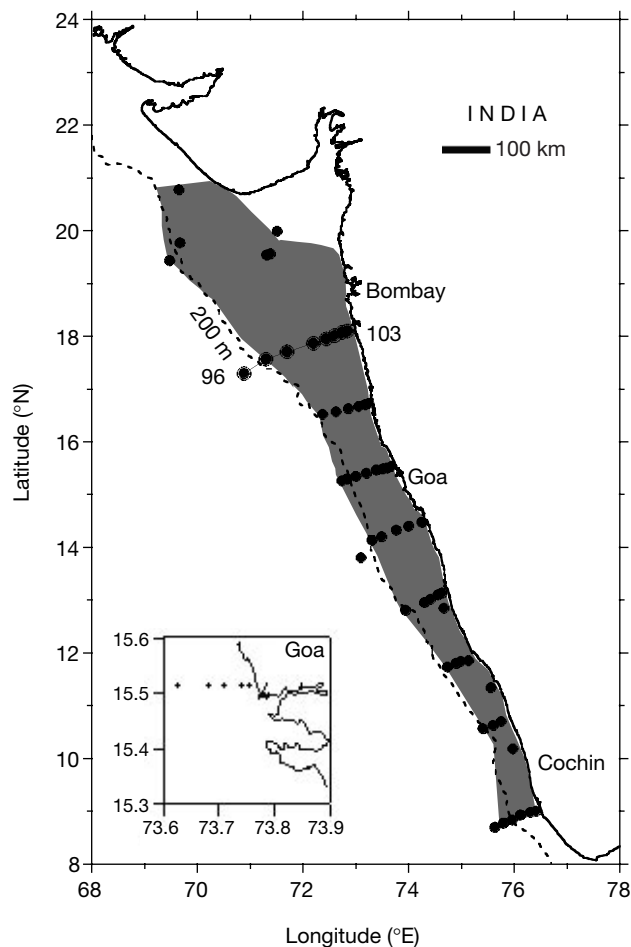


Figure 1 Zone of severe hypoxia on the western Indian shelf during September–October 1999 and locations of sampling sites. Zone of hypoxia is shown as shaded region ($O_2 < 0.5\text{ ml l}^{-1}$). Large circles represent stations occupied during the cruises whereas the five shallow stations (depth 6–28 m) comprising the shorter section off Goa are shown by small circles in the inset.

* Present addresses: Department of Geosciences, Princeton University, Princeton, New Jersey 08544-1003, USA (D.A.J.); and CEREGE, Université d'Aix-Marseille III, Europole de l'Arbois, BP 80, Cedex 4, 13545 Aix en Provence, France (V.V.S.S.).

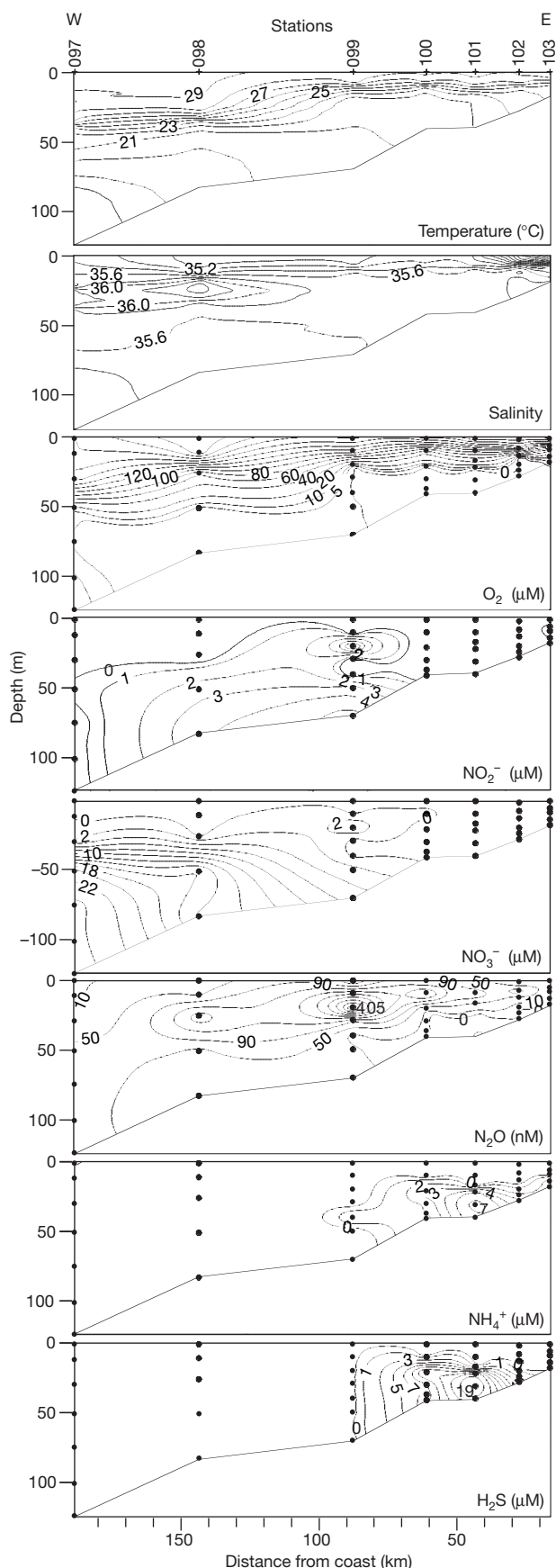


Figure 2 Distribution of properties along the cross-shelf transect off Bombay during 6–7 October 1999. See Fig. 1 for station locations.

and 14) are too limited to make any reasonable comparison with our results. However, under a UNDP/FAO project, O_2 data were collected¹⁵ along several cross-shelf transects between 7° N and 17° N during 1971–75. Near-bottom O_2 levels during these surveys were generally higher than those observed by us by over 10 μM . Although H_2S was not measured, ‘zero’ O_2 concentrations characteristic of sulphate-reducing systems were not recorded, and it is highly unlikely that, if present, H_2S would have gone undetected. Limited observations made subsequently (unpublished data available at the Indian National Oceanographic Data Centre), which also covered nutrients, again did not yield any ‘zero’ O_2 values, but frequent presence of NO_2^- in conjunction with low NO_3^- in sub-pycnocline waters indicated the occurrence of denitrification (see Supplementary Information). Therefore, we conclude that O_2 depletion over the western Indian shelf has intensified in recent years. At this point it cannot be established whether this change has been brought about by an increased nutrient loading or a modified physical forcing. As in other areas, consumption of nitrogen-based fertilizers in the coastal belt has increased by an order of magnitude over the past four decades. Surface $NO_3^- + NO_2^-$ concentrations in the Mandovi and Zuari estuaries (Goa) were as high as 10 μM during a period of

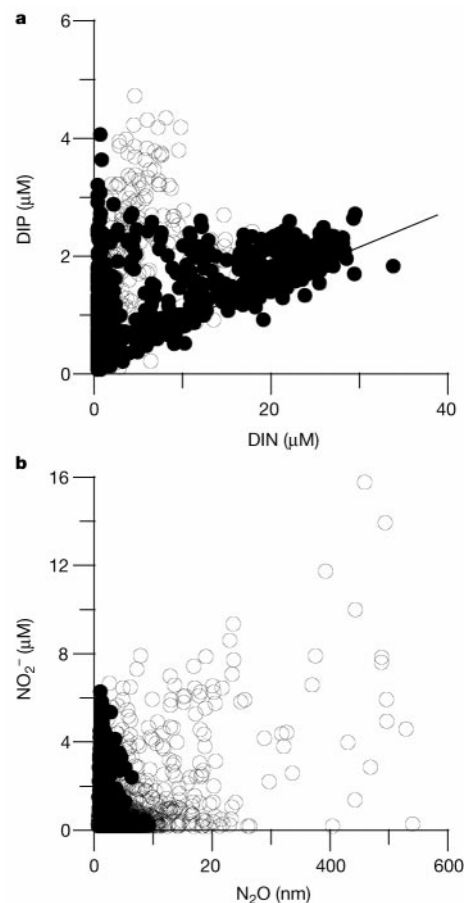


Figure 3 Relative changes in nutrients and N_2O in suboxic waters. **a**, Plot of dissolved inorganic phosphorus (DIP) versus dissolved inorganic nitrogen ($DIN = NO_3^- + NO_2^- + NH_4^+$ [ammonium]) for all samples taken from depths of more than 200 m (open circles denote samples where NH_4^+ concentration exceeded 0.5 μM). Data from an earlier cruise (SK103, 26.6.95–15.7.95) have also been included in this analysis. We note that the enhanced DIN concentration at DIP greater than $\sim 3 \mu\text{M}$ are due to NH_4^+ accumulation in anoxic waters. The straight line gives the Redfield relationship in oxic waters. **b**, Plot of NO_2^- versus N_2O in the shallow (open circles, present study) and deep (closed circles, unpublished data) suboxic zones.

heavy runoff in July 2000 whereas rainwater collected at the National Institute of Oceanography, Goa, contained on an average $4\text{ }\mu\text{M}$ of NO_3^- . However, although significant nitrogen loading occurs through both runoff and precipitation, elevated surface NO_3^- concentrations were not noticed on any of the cruises except when the pycnocline was eroded by turbulence. This indicates that the nitrogen added with fresh water was quickly removed by algae.

All observations made during the upwelling period indicated intense denitrification over the inner- and mid-shelf north of 12°N as evident from the relationship between dissolved inorganic nitrogen (DIN) and dissolved inorganic phosphorus (DIP) for the pooled data (Fig. 3a). These data fall into two groups—one exhibiting the Redfield trend commonly seen in oxygenated waters with a DIN to DIP slope close to the theoretical value of 16 (ref. 16), and the other the denitrification–sulphate reduction trend characterized by low DIN but high DIP in reducing waters. In many of the samples belonging to the latter group, N_2O accumulated to levels previously not seen in sea water (Figs 2 and 3b). The highest N_2O concentration reported so far is 173 nM from the coastal waters off Peru¹⁷; we measured scores of values exceeding this quantity, with a maximum of 533 nM (Fig. 3b). With the exception of a few surface data (maximum 436 nM ; $8,250\%$ saturation), the highest N_2O concentrations occurred in suboxic waters with extremely high NO_2^- levels, in contrast to the trend^{17–19} seen in the open ocean suboxic zones where NO_2^- build-up is almost invariably associated with a depletion of N_2O . Low N_2O concentrations in NO_2^- -bearing waters were also found, especially close to the bottom in the mid-shelf region, which resulted in a mid-depth N_2O maximum (Fig. 2). However, high N_2O levels were not confined to the mid-shelf, and extended to the inner-shelf (depth $< 30\text{ m}$) during periods when sulphate reduction did not occur. Near-bottom depletion of N_2O was noticed less frequently in this zone and the concentrations varied greatly with time (see Supplementary Information).

Results of incubation of a sample with initial O_2 , NO_3^- and NO_2^- contents of 15 , 21.8 and $0.06\text{ }\mu\text{M}$, respectively, showed that NO_3^- was largely converted to NO_2^- in the first two days, and the combined $\text{NO}_3^- + \text{NO}_2^-$ concentration rapidly fell to nearly zero in six days (Fig. 4). The N_2O concentration rose from an initial value of 48.5 nM to a maximum of $1,568\text{ nM}$ before declining sharply after five days. All oxidized forms were completely consumed in eight days.

The transient accumulation of N_2O during incubation and the coincidence of high N_2O and NO_2^- levels indicate that denitrification is probably the major process responsible for the unusual N_2O accumulation in our study area, even though the open-system field

data may differ somewhat from the pattern seen in the closed-system incubation setup. For example, the large mid-depth build-up of N_2O at Station 99 (Fig. 2) was associated with NO_3^- and NO_2^- concentrations of 4.0 and $6.4\text{ }\mu\text{M}$, respectively, whereas in the incubation experiment N_2O accumulation occurred only after NO_3^- had been largely depleted. On the other hand, the highest N_2O concentration (533 nM) came from a sample that had lost virtually all of its NO_3^- and NO_2^- . However, the low O_2 conditions are also known to enhance the yield of N_2O during nitrification²⁰, and so the possibility of some contribution also coming through this pathway cannot be excluded.

A very similar pattern of N_2O accumulation has been reported from some freshwater lakes and attributed to a high yield of N_2O during denitrification in the presence of small amounts ($\sim 5\text{ }\mu\text{M}$) of O_2 (ref. 21). We propose that such an accumulation may be a common feature of all aquatic bodies where suboxia occurs close to the surface. The O_2 concentration can affect N_2O production by regulating (1) nitrification activity and the yield of N_2O during this process; (2) rate of NO_3^- production and hence its availability for denitrification (the low N_2O concentration in near-bottom waters may in part be due to this effect); and (3) N_2O reductase activity. It has been reported that in a young denitrifying system the consumption of N_2O by denitrifiers occurs only after *de novo* synthesis of N_2O reductase²². Thus, frequent aeration through turbulence can rejuvenate suboxic waters (affecting the activity of this enzyme) when denitrification occurs at shallow depths, thereby sustaining a very high N_2O yield. Moreover, N_2O so produced can easily escape to the atmosphere. Such sites are therefore expected to make a disproportionately large contribution to the emission of N_2O to the atmosphere.

The high rate of N_2O production over the western Indian shelf sustains very high N_2O concentrations at the surface (5.3 – 436 nM , average 37.6 nM , $n = 216$). Using this average value and employing two different models of sea to air fluxes^{23,24} the flux of N_2O to the atmosphere has been computed as 40 – $268\text{ }\mu\text{mol m}^{-2}\text{ d}^{-1}$ at wind speeds of 5 – 10 m s^{-1} . Since high surface N_2O levels prevail even during the early phase of upwelling²⁵, we extrapolated this rate over a period of six months and an area of $180,000\text{ km}^2$ to arrive at an overall N_2O efflux of 0.06 – 0.39 Tg of N_2O from our study region. This is roughly of the same magnitude as the most recent estimate of N_2O efflux (0.39 Tg of $\text{N}_2\text{O yr}^{-1}$) from the entire Arabian Sea²⁶. It may be pointed out that this estimate should be regarded as a minimum value because a significant fraction of N_2O would be transported out of the shelf and be ventilated to the atmosphere elsewhere.

If the O_2 depletion over the western Indian shelf has intensified in recent years, as may be indicated by our data, then the N_2O flux computed above should have a substantial human-induced component that is expected to increase further in future. Apart from the enhanced nutrient loading, the summer monsoon that causes both upwelling and freshwater influx into the coastal zone may intensify in response to global warming²⁷. A similar effect has been predicted for the hypoxic zone off Louisiana²⁸ where, as in other similar systems such as the Black Sea and the Baltic, the nitrogen loading and its transformation rates are expected to be higher. Additional hypoxic sites may develop in areas such as the Bay of Bengal. Although an enhanced rate of denitrification in the rapidly expanding shallow hypoxic zones is unlikely to affect the oceanic nitrogen inventory very much (it will reflux part of the anthropogenic inputs to the ocean), its impact on the global N_2O budget may be potentially serious; it deserves to be evaluated and considered in projections of future build-up of greenhouse gases in the atmosphere. □

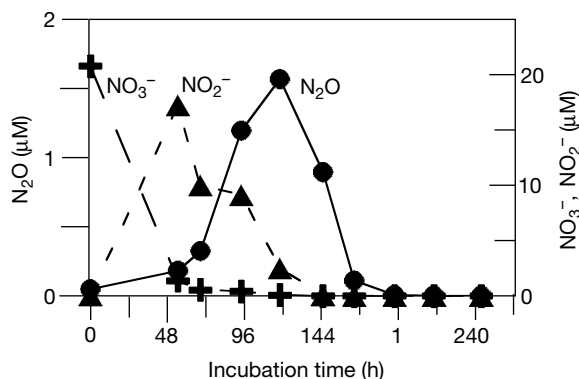


Figure 4 Changes in concentrations of dissolved nitrogen species with time in an incubation experiment. This sample was collected from 40 m at $12^\circ 54.9'\text{N}$, $74^\circ 18.0'\text{E}$ (station depth 64 m) on 30 August 1997.

Methods

Five cruises were undertaken aboard the research vessels *Sagar Sampada* (SS) and *Sagar Kanya* (SK): SS158 (22.8.97–2.9.97), SK137 (20.7.98–17.8.98), SK138a (30.9.98–4.10.98), SK148 (4.9.99–10.10.99) and SK149d (3.12.99–9.12.99). Sampling locations (Fig. 1) were

more or less the same on all cruises except that observations north of Goa were made only on SK148 and SK149d, and only the cross-shelf transect off Goa was worked during SK138a. The shallower transect off Goa was regularly sampled (on 15 occasions) during August–December for three years (1997–1999).

Water samples during the cruises were collected using Seabird CTD (conductivity–temperature–depth) rosette systems, and hydrocasts were made with Niskin bottles for observations in shallow waters off Goa. O_2 and H_2S analyses were performed by titrimetric (Winkler) and colorimetric (methylene blue) techniques, respectively²⁹, within three hours of collection. Nutrients were analysed using an autoanalyser²⁹. N_2O concentration was determined²⁵ by head-space extraction with helium followed by injection into a Hewlett-Packard 5890 Series II gas chromatograph equipped with an electron capture detector (precision ~4%). Percentage saturation was computed with reference to the solubilities given in ref. 30. For incubation experiments subsamples from a 20-litre Go-flo bottle were transferred into polyethylene-lined, airtight bags (10 litre), previously flushed with helium, avoiding any atmospheric contamination and incubated in the dark on board ship close to the *in situ* temperature. Aliquots were analysed periodically for NO_3^- , NO_2^- and N_2O .

Received 11 April; accepted 30 September 2000.

1. Turner, R. E. & Rabalais, N. N. Coastal eutrophication near the Mississippi River delta. *Nature* **368**, 619–621 (1994).
2. Diaz, R. J. & Rosenberg, R. Marine benthic hypoxia: A review of its ecological effects and the behavioural responses of benthic macrofauna. *Oceanogr. Mar. Biol. Annu. Rev.* **33**, 245–303 (1995).
3. Rabalais, N. N. Oxygen depletion in coastal waters. [online] (cited 3/3/2000) (http://state_of_coast.noaa.gov/bulletins/html/hyp_09/hyp.html) (NOAA State of the Coast Report, National Oceanic and Atmospheric Administration (NOAA), Silver Spring, Maryland, 2000).
4. Malakoff, D. Death by suffocation in the Gulf of Mexico. *Science* **281**, 190–192 (1998).
5. Houghton, J. T. *et al.* (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, Cambridge, 1996).
6. Banse, K. On upwelling and bottom-trawling off the southwest coast of India. *J. Mar. Biol. Ass. India* **1**, 33–49 (1959).
7. Banse, K. Hydrography of the Arabian sea shelf of India and Pakistan and effects on demersal fishes. *Deep-Sea Res.* **15**, 45–79 (1968).
8. Shetye, S. R. *et al.* Hydrography and circulation off the west coast of India during the Southwest Monsoon 1987. *J. Mar. Res.* **48**, 359–378 (1990).
9. Jayakumar, D. A., Naqvi, S. W. A., Narvekar, P. V. & George, M. D. Methane in coastal and offshore waters of the Arabian Sea. *Mar. Chem.* (in the press).
10. Calvert, S. E. & Price, N. B. Upwelling and nutrient regeneration in the Benguela Current, October, 1968. *Deep-Sea Res.* **18**, 505–523 (1971).
11. Codispoti, L. A. & Packard, T. T. Denitrification rates in the eastern tropical South Pacific. *J. Mar. Res.* **38**, 453–477 (1980).
12. Naqvi, S. W. A. Some aspects of the oxygen-deficient conditions and denitrification in the Arabian Sea. *J. Mar. Res.* **45**, 1049–1072 (1987).
13. Codispoti, L. A. *et al.* High nitrite levels off northern Peru: A signal of instability in the marine denitrification rate. *Science* **233**, 1200–1202 (1986).
14. Carruthers, J. N., Gogate, S. S., Naidu, J. R. & Laevastu, T. Shoreward upslope of the layer of minimum oxygen off Bombay: Its influence on marine biology, especially fisheries. *Nature* **183**, 1084–1087 (1959).
15. Progress Report No. 3, 1–13 (UNDP/FAO Pelagic Fishery Project, Bergen/Cochin, 1973).
16. Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, M. N.), Vol. 2, 26–77 (Interscience, New York, 1963).
17. Codispoti, L. A. *et al.* in *Oceanography of the Indian Ocean* (ed. Desai, B. N.) 271–284 (Oxford-IBH, New Delhi, 1992).
18. Law, C. S. & Owens, N. J. P. Significant flux of atmospheric nitrous oxide from the northwest Indian Ocean. *Nature* **346**, 826–829 (1990).
19. Naqvi, S. W. A. & Noronha, R. J. Nitrous oxide in the Arabian Sea. *Deep-Sea Res.* **38**, 871–890 (1991).
20. Goreau, T. J. *et al.* Production of NO_2^- and N_2O by nitrifying bacteria at reduced concentrations of oxygen. *Appl. Environ. Microbiol.* **40**, 526–532 (1980).
21. Yoh, M., Terai, H. & Saijo, Y. Accumulation of nitrous oxide in the oxygen deficient layer of freshwater lakes. *Nature* **301**, 327–329 (1983).
22. Firestone, M. K. & Tiedje, J. M. Temporal change in nitrous oxide and dinitrogen from denitrification following onset of anaerobiosis. *Appl. Environ. Microbiol.* **38**, 673–679 (1979).
23. Liss, P. & Merlivat, L. in *The Role Of Air-Sea Exchange In Geochemical Cycling* (ed. Buat-Menart, P.) 113–128 (Riedel, Dordrecht, 1986).
24. Wanninkhof, R. H. Relationship between wind speed and gas exchange over the ocean. *J. Geophys. Res.* **97**, 7373–7382 (1992).
25. Naqvi, S. W. A. *et al.* Budgetary and biogeochemical implications of N_2O isotope signatures in the Arabian Sea. *Nature* **394**, 462–464 (1998).
26. Bange, H. W. *et al.* A revised nitrogen budget for the Arabian Sea. *Global Biogeochem. Cycles* (in the press).
27. Kitoh, A., Yukimoto, S., Noda, A. & Motoi, T. Simulated changes in the Asian summer monsoon at times of increased atmospheric CO_2 . *J. Meteorol. Soc. Japan* **75**, 1019–1031 (1997).
28. Justić, D., Rabalais, N. N. & Turner, R. E. Effects of climate change on hypoxia in coastal waters: a doubled CO_2 scenario for the northern Gulf of Mexico. *Limnol. Oceanogr.* **41**, 992–1003 (1996).
29. Grasshoff, K., Ehrhardt, M. & Kremling, K. (eds) *Methods of Seawater Analysis* (Verlag Chemie, Weinheim, 1983).
30. Weiss, R. F. & Price, B. A. Nitrous oxide solubility in water and seawater. *Mar. Chem.* **8**, 347–359 (1980).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work forms a part of the LOICZ-India programme of the Department of Ocean Development (DOD); we thank the Sagar Sampada and Sagar Kanya Management Cells of

DOD for the generous allocation of ship time. The custom-made bags used for incubation experiments were given by B. Ward and J. Fernandes, B. Jose and D. Shenoy extended technical assistance at sea. Discussions with K. Banse, L. Codispoti and N. Rabalais greatly improved the content of the paper.

Correspondence and requests for materials should be addressed to S.W.A.N. (e-mail: naqvi@csnio.ren.nic.in).

Microseismological evidence for a changing wave climate in the northeast Atlantic Ocean

I. Grevemeyer*, R. Herber† & H.-H. Essen‡

* Department of Earth Sciences, University of Bremen, Klagenfurter Str., Bldg. GEO, 28359 Bremen, Germany

† Geophysical Observatory, University of Hamburg, Kuhtrift 18, 21075 Hamburg, Germany

‡ Institute of Oceanography, University of Hamburg, Troplowitzstr. 7, 22529 Hamburg, Germany

One possible consequence of a change in climate over the past several decades is an increase in wave heights, potentially threatening coastal areas as well as the marine industry^{1–4}. But the difficulties in observing wave heights exacerbates a general problem of climate-change detection: inhomogeneities in long-term observational records owing to changes in the instruments or techniques used, which may cause artificial trends^{5,6}. Ground movements with periods of 4–16 seconds, known as microseisms, are associated with ocean waves and coastal surf^{7–10}, and have been recorded continuously since the early days of seismology. Here we use such a 40-year record of wintertime microseisms from Hamburg, Germany, to reconstruct the wave climate in the northeast Atlantic Ocean. For the period 1954–77, we detect an average of seven days per month with strong microseismic activity, without a significant trend. This number increases significantly in the second half of the record, reaching approximately 14 days of strong microseisms per month. The implied increase in northeast Atlantic wave height over the past 20 years parallels increased surface air temperatures¹¹ and storminess¹² in this region, suggesting a common forcing.

Observations of the Earth's near-surface temperature show a global increase since 1901, occurring from 1925–1944 and 1978–1997. Over these periods global temperature rose by 0.37 and 0.32 K, respectively¹¹. The temperature change over the past decades is unlikely to be entirely due to internal climate variability^{13–15} and has been attributed to changes in the concentration of greenhouse gases caused by human activity^{16,17}. There are, however, basic physical relationships between temperature, air pressure and wind fields^{17,18} (and hence wave fields^{4,19}). Consequently, many people are concerned about the possibility of an intensification of extratropical storms^{20,21}. However, because of inhomogeneities in historical data sets, the impact of climate changes on the temporal evolution of the storm and wave climate is difficult to reveal.

In climatology, homogeneity and therefore quality of data is essential^{5,6}. A climatological time series is termed homogeneous²² if the variations exhibited by the series are solely the result of the vagaries of the weather and climate. Therefore, the methodological challenge with the analysis of historical data sets is the discrimination between signals that reflect real changes and signals that reflect changes that are attributable to improved instrumental accuracies, altered environmental conditions, observational practices, data

coverage and analysis routines. In terms of wave height, data are available from reports of visual assessments from ships of opportunity and lighthouses, from river buoys and shipborne wave recorders at ocean weather stations; wave height maps have also been constructed for the purpose of ship routing from wind analyses. Analyses of these data have revealed a substantial worsening of the wave climate in the north Atlantic Ocean^{1–3}. However, these data are sparse and suffer from various inhomogeneities²³.

In the late nineteenth century seismologists started to record movements of the ground continuously. In addition to ordinary earthquakes seismographs almost always record small movements in the Earth's crust, called microseisms. It has rapidly been recognized that several parts of microseisms (periods of 4 to 16 s) are strongly correlated with oceanographic forcing and atmospheric disturbances. In 1904 Wiechert⁷ considered the surf to be the main

cause of microseisms produced by extratropical low-pressure areas. Many authors confirmed this idea by experimental data^{8,24,25}. However, microseismic energy could also be generated in the oceans away from coastal areas^{9,26,27} and Gutenberg⁹ showed that microseisms could be used in weather forecasting, especially in locating tropical disturbances. Later a mathematical theory was established^{10,27} that explains the generation of microseisms by standing ocean waves and coastal surf. Although this permitted ocean wave heights to be calculated from seismometer data²⁸, most seismologists considered microseisms primarily to be noise in seismic recordings. Today, however, these data can be re-examined to assess the wave climate of the twentieth century.

We have shown that microseisms at the seismological station HAM (Hamburg, Germany) are related to the northeast Atlantic wave climate²⁴. Large-amplitude microseismic ground motions occurring at HAM are primarily related to secondary microseisms^{10,26,27} with periods of 6–8 s. Here, we present historical microseismic data recorded between 1954 and 1998. This period covers the time where other climatological data sets detected major changes in, for example, the surface air temperature¹¹, the winter-time north Atlantic oscillation (NAO) index¹⁸ and the northwestern European storm climate^{4,12}. The seismic records used are those from a Wiechert vertical seismograph (1,300 kg pendulum) and a Sprengnether long-period vertical seismometer. The instruments were operated from 1954–1975 and 1974–1998, respectively. To yield absolute ground motion from different seismometers the recordings have to be convolved with an instrument-dependent response function. Historical seismological data, however, are analogue recordings. In the analysis of microseisms it is not feasible to digitize 40 years of continuously recorded data. The standard approach to obtaining daily statistics of microseisms from analogue data is to read the amplitude and frequency of the strongest microseismic ground motions over a fixed time window of 1–2 hours per day²⁹. Using digital data from the wintertime microseisms of the years 1992 to 1998 (ref. 24) we were able to show that this technique provided reliable averages. In addition, we introduce a measure called the microseismic index that should be independent

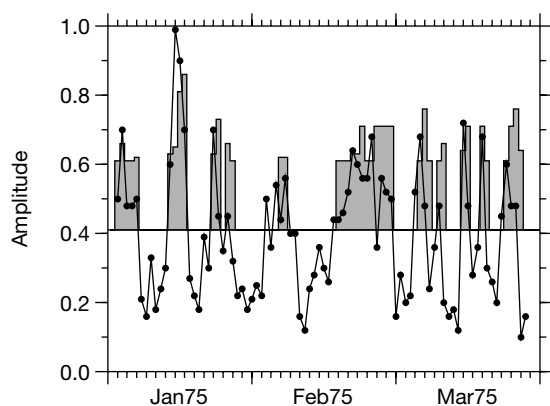


Figure 1 Correlation between daily microseism data (periods 6–8 s) from an old Wiechert pendulum and a modern Sprengnether seismometer for January–March 1975. Grey shaded areas indicate days with significant microseismic ground motions derived from the Sprengnether seismometer. The threshold amplitude for the Wiechert pendulum was changed until both instruments provided a common trend.

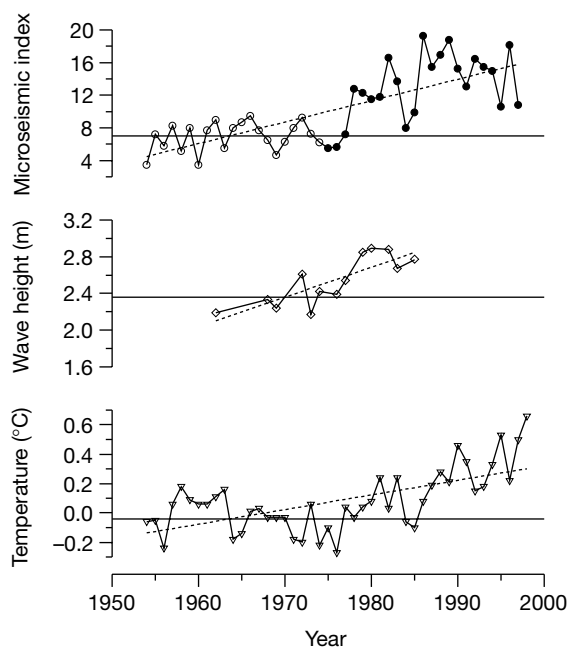


Figure 2 Time series of the wintertime microseisms recorded at the station HAM (Hamburg, Germany) over the last four decades. Microseisms at HAM are produced by waves in the northeast Atlantic and coastal surf along northern Europe's coastlines^{8,10,24}; thus, they could be used to assess changes of the wave climate. Solid dots are from a

Sprengnether seismometer and open dots are from a Wiechert seismometer. Also shown are the mean annual significant wave heights at the Seven Stones Light Vessel off Land's End¹ and the Northern Hemisphere temperature anomaly¹¹. Horizontal lines are the 1954–1977 averages and broken lines indicate linear trends revealed by the time series.

of the instrumental response within the narrow band of periods considered in our analysis. To characterize days with strong microseisms, we define a threshold amplitude of ground movements which is significantly above the noise level and above the average microseisms recorded during summer seasons. If this threshold is reached, we consider the day to be affected by microseismic activity. Thus, the microseismic index defines the number of days per month affected by strong microseisms and is therefore more a qualitative measure indicating overall changes of the wave climate rather than a quantitative measure yielding absolute changes of wave height.

The threshold amplitude itself might be in some way arbitrary. Another value, however, merely produces a linear shift of the time series, without affecting the overall trend revealed by the data. To correlate and normalize recordings from the Sprengnether and the Wiechert seismometer we used the daily measurements and changed the threshold amplitude for the Wiechert until both instruments supported a common trend during the time of overlap. Within the relatively narrow frequency band of secondary microseisms, any bias due to differences in the response of the seismometers should be minimized. This is demonstrated by the excellent correlation between the two time series (Fig. 1). This approach provided a data set covering 40 years which has no substantial inhomogeneities. However, because most storms and the resulting strong microseisms are generally restricted to the wintertime²⁹, we only rely on the months October to March.

Figure 2 shows the annual averages of the microseismic index. Like other climatological time series it displays a considerable degree of year-to-year variability. There is a noticeable increase (estimated by a linear trend calculated using least squares) in the frequency of microseismic storms, that is, the number of days per month affected by microseisms increases by 0.26 per year. However, the data clearly indicate two different periods; from 1954–1977 the number of microseisms that occur increases only slightly, settling at a steady rate of 7 days per month, but increases significantly from 1978–1998 to 14 days per month. This fact is visible even in the data obtained from the Sprengnether seismometer alone; operated since 1974, it provided the lowest values in its first years. Thereafter, values double within a few years.

The same trend could be observed at the Seven Stones Light Vessel off Land's End¹ (Fig. 2). It was the first site in the world where a wave recorder was installed; annual mean values of significant wave height were used¹ to examine whether the northeast Atlantic has become rougher in recent years. The measurements published cover the years 1962 to 1985. We note that, just as for the microseismic index, annual means obtained before 1978 are always lower than those obtained thereafter. Unfortunately, there is no homogeneous data set available for comparison. However, problems of inhomogeneities were overcome by feeding a numerical wave model (WAM)¹⁹ with historical surface pressure and wind distribution data. This simulation¹⁹ suggested, for the Norwegian Sea at the weather station OWS Mike, an increase of significant wave heights over the period 1955–1994 of the order of 8 cm yr⁻¹ for the annual maxima. The increase in wave height was even more significant in the second part of the simulated period, 1975–1994. During that time the maxima of significant wave height increased by more than 17 cm yr⁻¹. A formal correlation of microseisms and wave hindcast data gave a regression coefficient of $r = 0.4$ which increases to 0.61 when using data smoothed with a five-year gaussian filter. Even the 99 and 90 percentiles indicate similar trends. In general, however, a two-step evolution is not evident in the simulations of the WASA project. Nonetheless, the model output suggests that statistics of significant wave height in the northern North Sea and the Norwegian Sea have undergone a steady increase since 1954 (refs 4, 19). Thus, in addition to previous studies, microseisms at HAM present a new and homogeneous data set that suggests and supports a worsening of the northeast Atlantic wave climate over the last two decades.

If we compare microseisms with the Northern Hemisphere (or global) temperatures (Fig. 2) we observe a similar trend; temperatures remaining nearly at a constant level between 1954 and 1977 and a warming of 0.32 K from 1978–1997 (ref. 11). Similar trends are also revealed by the NAO index and the northwestern European storm climate¹². A two-phase evolution, however, is not evident; but in terms of an overall increase the correlation between the time series is statistically significant. The similarities between the different climatological records strongly suggest a common forcing. Recent simulations and analyses of the Earth's temperature pattern exclude purely natural forcing and attribute it largely to changes in the concentration of greenhouse gases and aerosol loading due to human activity^{13,17}. Therefore, it seems reasonable to propose that greenhouse forcing affects the ocean's wave climate and hence coastal surf and storm surges along northern Europe's coastlines, which in turn produced the observed increase of microseisms. However, large ocean waves and hence significant microseisms are generally related to the very high wind speed of storms. The storm climate itself seems to be comparable with that at the beginning of the twentieth century^{4,12}, though. Work is now required to backtrack microseisms and hence the wave climate into the early twentieth and late nineteenth century. Such a time series will allow us to understand the interaction between the NAO index, surface air temperature, storm frequency and intensity and the north Atlantic wave climate on a longer timescale. □

Received 23 February; accepted 24 August 2000.

- Carter, D. J. T. & Draper, L. Has the north-east Atlantic become rougher? *Nature* **332**, 494 (1988).
- Bouws, E., Jannick, D. & Komen, G. J. On increasing wave height in the North Atlantic Ocean. *Bull. Am. Meteorol. Soc.* **77**, 2275–2277 (1996).
- Bijl, W. Impact of wind climate change on the surge in the southern North Sea. *Clim. Res.* **8**, 45–59 (1997).
- WASA Group. Changing waves and storms in the Northeast Atlantic? *Bull. Am. Meteorol. Soc.* **79**, 741–760 (1998).
- Karl, T. R., Quayle, R. G. & Groisman, P. Y. Detecting climate variations and changes: new challenges for observing and data management systems. *J. Clim.* **6**, 1481–1494 (1993).
- Jones, P. D. in *Analysis of Climate Variability* (eds von Storch, H. & Navarra, A.) 53–76 (Springer, Berlin, 1995).
- Wiechert, E. Discussion, Verhandlung der zweiten Internationalen Seismologischen Konferenz, Strasbourg. *Gerlands Beitr. Geophysik* **2**, 41–43 (1904).
- Gutenberg, B. Untersuchungen über die Bodenunruhe mit Perioden von 4–10 Sekunden in Europa. *Veröff. Zentr. Bur. Int. Seismol. Assoc.* **106** (1921).
- Gutenberg, B. Microseisms and weather forecasting. *J. Meteorol.* **4**, 21–28 (1947).
- Hasselmann, K. Statistical analysis of the generation of microseisms. *Rev. Geophys.* **1**, 177–210 (1963).
- Jones, P. D., New, M., Parker, D. E., Martin, S. & Rigor, I. G. Surface air temperature and its changes over the past 150 years. *Rev. Geophys.* **37**, 173–199 (1999).
- Alexandersson, H., Smith, T., Iden, K. & Tuomenvirta, H. Long-term trend variations of the storm climate over NW Europe. *Glob. Atmos. Ocean Sys.* **6**, 97–120 (1998).
- Stouffer, R. J., Manabe, S. & Vinnikov, K. Y. Model assessment of the role of natural variability in recent global warming. *Nature* **367**, 634–636 (1994).
- Santer, B. D. *et al.* A search for human influences on the thermal structure of the atmosphere. *Nature* **382**, 39–45 (1996).
- Mann, M. E., Bradley, R. S. & Huges, M. K. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature* **392**, 779–787 (1998).
- Tett, S. F. B., Stott, P. A., Allen, M. R., Ingram, W. J. I. & Mitchell, J. F. B. Causes of twentieth-century temperature change near the Earth's surface. *Nature* **399**, 569–572 (1999).
- Shindell, D. T., Miller, R. L., Schmidt, G. A. & Pandolfo, L. Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature* **399**, 452–455 (1999).
- Rodwell, M. J., Rowell, D. P. & Folland, C. K. Oceanic forcing of the wintertime North Atlantic oscillation and European climate. *Nature* **399**, 320–323 (1999).
- Günther, H. *et al.* The wave climate of the Northeast Atlantic over the period 1955–1994: the WASA wave hindcast. *Glob. Atmos. Ocean Sys.* **6**, 121–163 (1998).
- Berz, G. Global warming and the insurance industry. *Interdisciplinary Sci. Rev.* **18**, 120–125 (1993).
- Berz, G. & Conrad, K. Stormy weather: the mounting wind-storm risk and consequences for the insurance industry. *Eurodecision* **12**, 65–69 (1994).
- Conrad, V. & Pollak, L. D. *Methods in Climatology* (Harvard Univ. Press, Cambridge, Massachusetts, 1962).
- WASA Group. Comment on "Increases in Wave Heights over the North Atlantic: a review of the evidence and some implications for the naval architect" by N. Hogben. *Trans. R. Inst. Naval Arch.* **137**, 107–110 (1994).
- Essen, H.-H., Klusmann, J., Herber, R. & Grevenmeyer, I. Do microseisms in Hamburg (Germany) reflect the wave climate of the North Atlantic? *Germ. J. Hydrogr.* **51**, 33–45 (1999).
- Darbyshire, J. Analysis of twenty microseism storms during the winter of 1987–1988 and comparison with wave hindcasts. *Phys. Earth Planet. Int.* **63**, 181–195 (1990).
- Longuet-Higgins, M. S. & Ursell, F. Sea waves and microseisms. *Nature* **162**, 700 (1948).
- Longuet-Higgins, M. S. A theory of the origin of microseisms. *Phil. Trans. R. Soc. Lond. A* **243**, 1–35 (1950).

28. Bromirski, P. D., Flick, R. E. & Graham, N. Ocean wave height determined from inland seismometer data: Implications for investigating wave climate changes in the NE Pacific. *J. Geophys. Res.* **104**, 20753–20766 (1999).
29. Bath, M. *An Investigation of the Uppsala Microseisms* (Institute of Meteorology, Royal Univ. Uppsala, Report No. 14, Uppsala, 1949).

Acknowledgements

We thank G. Spars for assistance in analysing the historical seismological records. This work benefited from support of the Deutsche Forschungsgemeinschaft for the SFB 512 "Cyclones and the North Atlantic Climate System".

Correspondence and requests for materials should be addressed to I.G. (e-mail: ingo@geophys2.uni-bremen.de).

Fine-scale genetic structuring on *Manacus manacus* leks

Lisa Shorey*, Stuart Piertney†, Jon Stone‡ & Jacob Höglund*

* Population Biology/EBC, Uppsala University, Norbyvägen 18D, SE-752 36, Uppsala, Sweden

† NERC Molecular Genetics in Ecology Initiative, Department of Zoology, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

‡ Animal Ecology/EBC, Uppsala University, Norbyvägen 18D, SE-752 36, Uppsala, Sweden

Leks have traditionally been considered as arenas where males compete to attract females and secure matings. Thus, direct fitness benefits mediated through competition between males to fertilize females have been considered to be the primary force driving the evolution of lekking behaviour^{1,2}. Inclusive fitness benefits mediated through kin selection³ may also be involved in lek formation and evolution^{4,5}, but to date this theory has been largely ignored. According to kin-selection theory, both reproducing and non-reproducing males may gain indirect inclusive fitness benefits. If females are attracted to larger leks, non-reproducing males add attractiveness to a lek, and therefore, in a genetically structured population, boost the reproductive success of kin. Theory predicts that the attractiveness of leks is plastic, and that males establish themselves on a lek in which the top male, in terms of reproductive success, is a close relative⁶. Here we show that in white-bearded manakins (*Manacus manacus*), for which larger leks are more attractive to females^{7,8} and so secure the maximum number of matings, there is extraordinary fine-scale genetic structure, with leks being composed of clusters of related kin. We propose that males establish themselves where they find relatives to such an extent that they form groups within leks, and that such behaviour is consistent with kin-selection theory to maximize reproductive success of the group.

Manacus manacus males aggregate on display grounds (leks) to attract females for the purpose of mating. Each male defends a small court on the lek where he performs an acrobatic display. Leks are present all year round, annual mortality is low and birds may live up to 15 years^{9,10}. Birds remain at the same court all year and from one year to the next. Among these courts, however, male mating success is highly skewed⁷.

It is assumed that all males on a lek aim to increase their individual fitness by fathering as many offspring as possible. There are, however, potential indirect inclusive fitness effects that could operate, which have largely been neglected as a cue for lek evolution. In many lekking species, females prefer to mate in larger male aggregations than in smaller aggregations or with single males¹, which for a number of species leads to the general positive relationship between lek size and number of mating females. This is

certainly the case in *M. manacus*^{7,8}, for which larger lek sizes appear to attract more female visits (Spearman's coefficient of rank correlation (r_s) = 0.9, n = 5, P = 0.037; data from leks within our study in 1999 and 2000 and from two leks in ref. 7 where comparable data exist, female visits corrected for observational effort). As such, an individual could enhance his own inclusive fitness by joining a lek in which a relative is dominant and increasing the relative's reproductive success^{4–6}.

We used allele frequencies at four polymorphic microsatellite loci to estimate pairwise relatedness between males on two leks. Relatedness information¹¹ was combined with data on positions of male courts to delimit groups of related individuals. *M. manacus* males positioned themselves on a lek with relatives (Fig. 1). Furthermore, they positioned themselves in the lek to form spatially separated clusters of relatives, each with one or more reproductively successful male(s) (Fig. 2).

M. manacus females choose centrally positioned males in better physiological condition (L.S., unpublished data). Relatives may be more likely to gather around a top male and be accepted if they help in the initial attraction of females to the lek. By clustering near a top male, a family member may also improve his chances of acquiring a central court in the event of the death or injury of his successful relative. As suggested by our data, there may be more than one such family group on a single lek.

Related males may end up on the same lek by chance, as has been suggested in black grouse (*Tetrao tetrix*), for which male kin association on leks may be due to limited natal dispersal of males⁴. By forming related groups within leks, however, it seems that establishment on *M. manacus* leks cannot simply be attributed to limited male natal dispersal where males join the lek closest to their place of birth. Once on the lek, active choice among residential and newly arrived birds must take place¹², as we observed clusters of relatives within leks. Our data show that one group on each lek (group 2 in each case) consisted of individuals with a high average relatedness consistent with first-order relationships, whereas the other two groups consisted of individuals with lower average relatedness (Fig. 1). This suggests that the dynamics of reproductive sharing and competition could vary among groups within leks. Notably, the group showing the lowest average pairwise relatedness (r) (group 1 on lek 2) showed a more equal sharing of matings among the males (Fig. 2).

M. manacus individuals are unlikely to learn the characteristics of

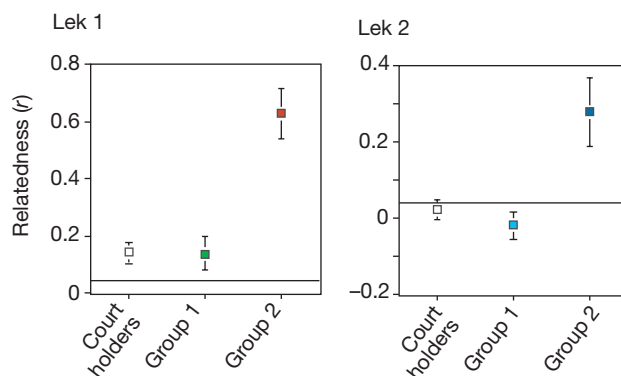


Figure 1 Mean relatedness values of male groupings. The line in each graph indicates the mean relatedness of both leks (bars indicate standard errors). Lek 1: green, group 1; red, group 2. Lek 2: light blue, group 1; dark blue, group 2. Statistical test, analysis of variance (ANOVA) in all cases. Lek 1: population versus court holders, $F_{1,2082} = 23.84$, $P < 0.0001$; population versus group 1, $F_{1,2082} = 3.95$, $P < 0.047$; population versus group 2, $F_{1,2088} = 153.3$, $P < 0.0001$. Lek 2: population versus court holders, $F_{1,2172} = 0.02$, not significant; population versus group 1, $F_{1,2094} = 1.13$, not significant; population versus group 2, $F_{1,2088} = 27.42$, $P < 0.0001$.

their relatives in the nest because clutch size is only one or two eggs and nest predation is high⁹. Thus, any male surviving to become an adult is unlikely to have shared the nest with any other bird in the population. The most plausible mechanism is that both prospecting floaters and established court holders use some kind of phenotype matching¹³ (probably self-referent phenotype matching) when juveniles float around the leks in the local population before settling on a lek.

Our study supports the hypothesis that kin selection is involved in lek evolution and supports a shift away from the model that leks are primarily arenas for intrasexual competition. Cooperation may also influence the evolutionary dynamics of leks, mediated either through kin selection or through by-product mutualism as suggested in long-tailed manakins (*Chiroxiphia linearis*)¹⁴. Moreover, our data suggest that wild populations that have previously been thought of as comprising unrelated individuals can be kin structured at a very fine level, and this structure can have profound effects on reproductive success. □

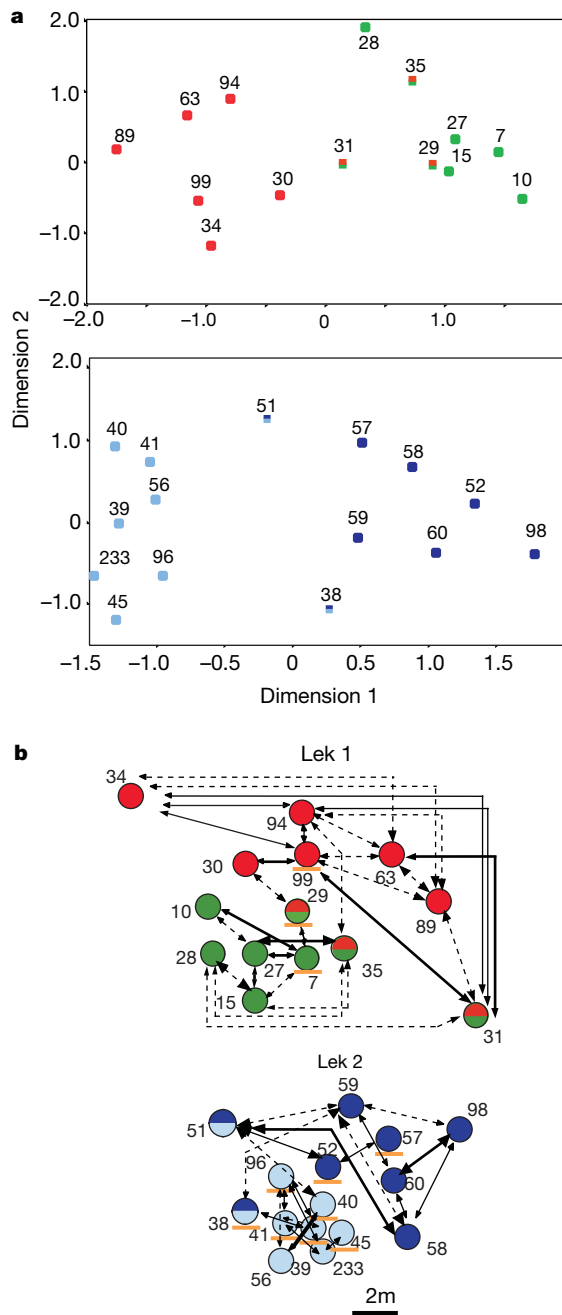


Figure 2 Levels of relatedness and spatial positioning on two *M. manacus* leks. **a**, Multidimensional scaling of the r values using a Euclidean distance model for leks 1 and 2. **b**, Maps showing the positions of individual court holders on each lek. Arrows connect related males: bold arrows, significant at $\alpha = 0.05$; thin arrows, $\alpha = 0.25$; dashed arrows, $r > 0.05$. Groups are coloured as in Fig. 1, males successful in obtaining copulations are underlined. We tested the spatial distribution of court holders by assigning them to one of either group found at each lek. Disregarding males spanning two groups, the observed spatial distribution of males was highly nonrandom (Fisher's exact test: lek 1, $P = 0.002$; lek 2, $P = 0.00058$). Males with two colours were related to both groups.

Methods

M. manacus leks were studied in the Paria area of Trinidad, West Indies ($10^{\circ}45'N$, $61^{\circ}15'W$). We caught the males ($n = 250$) on 10 leks during April 1998, using mist nets. Roughly 4 μ l of blood was taken from the wing vein and stored in EDTA buffer. Each bird was ringed with a unique colour combination and subsequently released to facilitate behavioural observation. We observed two leks (lek 1, $n = 34$; lek 2, $n = 27$) over an eight-week period to establish whether a male was a court holder, and if so his position in the lek and mating success. Fourteen and fifteen males defended constant territories in a lek, and were thus defined as court holders. The remaining individuals were occasional visitors or unestablished juveniles.

We extracted DNA from blood samples using standard proteinase K/phenol–chloroform procedures. Four microsatellite loci (Man3, Man7, Man6 and Man8) specifically cloned in *M. manacus* were genotyped, resolving 14, 9, 11 and 11 alleles for each locus, respectively. For details on primer sequences and laboratory procedures, see Supplementary Information. Briefly, alleles were amplified using polymerase chain reaction (PCR), separated using polyacrylamide electrophoresis and visualized using silver staining.

We obtained unambiguous genotypes for every bird at every locus. No locus was found to deviate from Hardy–Weinberg expectations. The allele frequencies were then used in a regression estimator¹¹ to estimate pairwise relatedness (r) among all individuals on the lek. We obtained 95% and 75% confidence intervals for each estimate by bootstrapping the loci 1,000 times for each pair of individuals (program obtainable by request from J.S.). We used multidimensional scaling (MDS) to provide an unbiased representation of kin groupings on leks. MDS transforms pairwise relatedness generated from the microsatellite data to distance between points in two-dimensional space. We then compared kin groupings with the permanent territory positions observed. Males were considered related if the estimated r was above 0.05, as the expected r based on all possible dyads for both leks was 0.04 ± 0.004 (s.e.). Individual pairwise relatedness values ranged from -0.188 to $+0.939$ (lek 1) and -0.219 to $+0.914$ (lek 2). Owing to the limited number of microsatellite loci studied only some dyads were significant at $\alpha = 0.05$ (α , type I error probability) (Fig. 2).

Received 22 May; accepted 21 August 2000.

- Höglund, J. & Alatalo, R. V. *Leks* (Princeton Univ. Press, Princeton, 1995).
- Wiley, R. H. Lekking in birds and mammals: Behavioral and evolutionary issues. *Adv. Study Behav.* **20**, 201–291 (1991).
- Hamilton, W. D. The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**, 1–16 (1964).
- Höglund, J., Alatalo, R. V., Lundberg, A., Rintamäki, P. T. & Lindell, J. Microsatellite markers reveal the potential for kin selection on black grouse leks. *Proc. R. Soc. Lond. B* **266**, 813–816 (1999).
- Petrie, M., Krupa, A. & Burke, T. Peacocks lek with relatives even in the absence of social and environmental cues. *Nature* **401**, 155–157 (1999).
- Kokko, H. & Lindström, J. Kin selection and the evolution of leks: whose success do young males maximise? *Proc. R. Soc. Lond. B* **263**, 919–923 (1996).
- Lill, A. Social organization and space utilization in the lek-forming white-bearded manakin, *M. manacus trinitatis* Hartert. *Z. Tierpsychol.* **36**, 513–530 (1974).
- Kokko, H., Sutherland, W. J., Lindström, J., Reynolds, J. D. & Mackenzie, A. Individual mating success, lek stability, and the neglected limitations of statistical power. *Anim. Behav.* **56**, 755–762 (1998).
- Snow, D. W. A field study of the black and white manakin, *Manacus manacus*, in Trinidad. *Zoologica* **47**, 65–104 (1962).
- Lill, A. Sexual behavior of the lek-forming white-bearded manakin (*Manacus manacus trinitatis* Hartert). *Z. Tierpsychol.* **36**, 1–36 (1974).
- Lynch, M. & Ritland, K. Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753–1766 (1999).
- Sherman, P. W. Birds of a feather lek together. *Nature* **401**, 119–120 (1999).
- Sherman, P. W., Reeve, H. K. & Pfennig, D. W. in *Behavioural Ecology An Evolutionary Approach* 4th edn (eds Krebs, J. R. & Davies, N. B.) 69–96 (Blackwell, Oxford, 1997).
- McDonald, D. B. & Potts, W. K. Cooperative display and relatedness among males in a lek-mating bird. *Science* **266**, 1030–1032 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Fitzjames, J. Johnston and T. Lievonen for help in the field, S. Griffith for advice and R. Dufva for help in the lab. The study was funded by the Swedish Natural Sciences Research Council (NFR) (to J.H.).

Correspondence and requests for materials should be addressed to L.S. (e-mail: lisa.shorey@ebc.uu.se).

Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants

Keith L. Adams*, Daniel O. Daley†, Yin-Long Qiu*‡, James Whelan† & Jeffrey D. Palmer*

* Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

† Biochemistry Department, University of Western Australia, Nedlands 6907, Australia

‡ Present address: Department of Biology, University of Massachusetts, Amherst, Massachusetts 01003, USA.

A central component of the endosymbiotic theory for the bacterial origin of the mitochondrion is that many of its genes were transferred to the nucleus. Most of this transfer occurred early in mitochondrial evolution¹; functional transfer of mitochondrial genes has ceased in animals². Although mitochondrial gene transfer continues to occur in plants³, no comprehensive study of the frequency and timing of transfers during plant evolution has been conducted. Here we report frequent loss (26 times) and transfer to the nucleus of the mitochondrial gene *rps10* among 277 diverse angiosperms. Characterization of nuclear *rps10* genes from 16 out of 26 loss lineages implies that many independent, RNA-mediated *rps10* transfers occurred during recent angiosperm evolution; each of the genes may represent a separate functional gene transfer. Thus, *rps10* has been transferred to the nucleus at a surprisingly high rate during angiosperm evolution. The structures of several nuclear *rps10* genes reveal diverse mechanisms by which transferred genes become activated, including parasitism of pre-existing nuclear genes for mitochondrial or cytoplasmic proteins, and activation without gain of a mitochondrial targeting sequence.

We surveyed total DNAs from 277 genera of angiosperms, representing 169 families, for the presence in mitochondrial DNA of the ribosomal protein gene *rps10*, using probes for its first exon and single intron in Southern blot hybridizations (Fig. 1). The gene was judged to be absent from mitochondrial DNA if there was no detectable (or, in two cases, severely diminished) hybridization with the *rps10* exon probe but there was good hybridization to the DNA in question using other mitochondrial gene probes. This inference is

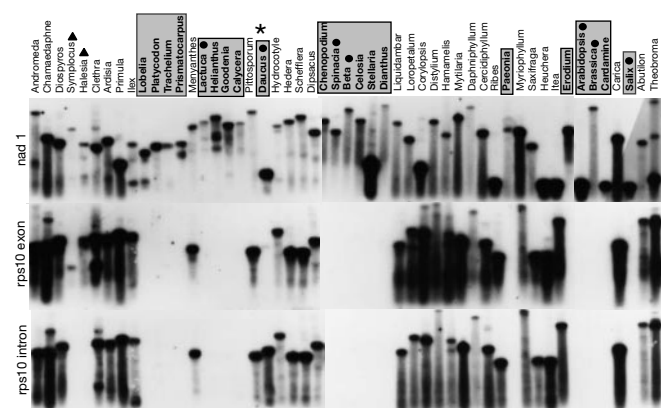


Figure 1 Southern blot hybridizations of 51 angiosperms (out of 277 examined in total) with the indicated angiosperm probes. Shading indicates taxa with no hybridization to the *rps10* exon probe. Triangles indicate no hybridization to only the *rps10* intron probe (a total of six intron losses were inferred); asterisk indicates no hybridization to only the exon probe. Bullets indicate species from which at least part of *rps10* was isolated from the nucleus.

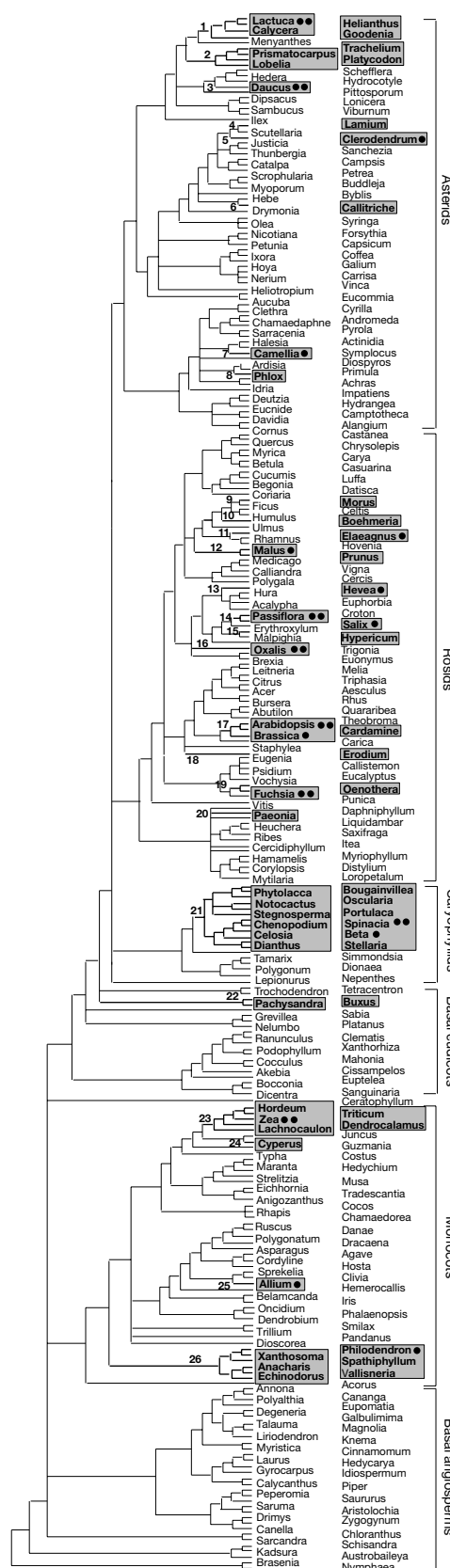


Figure 2 Sporadic loss of the mitochondrial *rps10* gene among the surveyed angiosperms. Shading and single bullets are the same as in Fig. 1; double bullets indicate those plants from which nuclear *rps10* was characterized in detail (see Figs 3–5). The 26 inferred losses of mitochondrial *rps10* are numbered; bold branches highlight these 26 loss lineages. Plant names beside the tree are staggered in two columns. The phylogenetic tree is based largely on the strict consensus tree from ref. 27, although other studies were consulted (see <http://www.bio.indiana.edu/~palmerlab>).

based on the hundreds-fold higher copy number of mitochondrial genomes relative to that of nuclear genomes in plants. Such a broad-scale survey was feasible because of the very low rate of nucleotide substitutions in plant mitochondrial genomes^{4,5}. Twenty-six separate losses of the *rps10* exon were inferred when the hybridization results were plotted on a phylogenetic tree of the surveyed angiosperms (Fig. 2). The losses are broadly and sporadically distributed among eudicots and monocots, but no losses were detected in basal angiosperms.

Putative nuclear *rps10* sequences were isolated by polymerase chain reaction (PCR) from species representing 16 loss lineages (Fig. 2; see Supplementary Information). Each sequence has two features diagnostic of a gene transferred to the nucleus: extensive divergence owing to the much higher nuclear mutation rate^{4,5}, and signatures of a processed RNA intermediate (lack of the group II intron present in virtually all mitochondrial *rps10* genes (Fig. 1), and T instead of C at sites of mitochondrial RNA editing). The 16 nuclear *rps10* genes could result from many (up to 16) separate transfers to the nucleus, or from as few as a single transfer in a common ancestor of monocots and eudicots. To infer the number and timing of *rps10* transfers, we further characterized several nuclear *rps10* genes. Many of the nuclear *rps10* genes have distinctive gene structures, notably fusions with pre-existing nuclear genes, that provide clues to the number of transfers and mechanisms of gene activation (that is, gain of sequences conferring nuclear expression and mitochondrial targeting).

Nuclear *rps10* from *Fuchsia* has a mitochondrial targeting sequence (presequence) that is homologous to, and shares two intron positions with, the presequence of the anciently transferred nuclear gene for the mitochondrial protein HSP70 (Fig. 3a). Thus, nuclear *rps10* in *Fuchsia* appears to have been activated by gaining a presequence, and possibly upstream regulatory elements, from *hsp70*. A similar case of a chimaeric transferred gene is illustrated

by *rps11* in rice⁶. A variation on presequence acquisition is exhibited by *rps10* from carrot, whose 5' and 3' ends are derived from the nuclear gene for the mitochondrial protein HSP22 (Fig. 3b). Much of the amino terminus of the carrot RPS10 fusion protein is cleaved on import into isolated mitochondria (Fig. 4). Thus, in carrot *rps10* has inserted into and parasitized a copy of *hsp22*, thereby gaining a presequence and possibly *cis*-regulatory elements. Another variation on this theme is shown by the transferred *rps14* gene in grasses, which was inserted into the intron of a pre-existing nuclear gene and is co-expressed with it through alternative splicing^{7,8}. Unlike the clearly identifiable presequences of *rps10* in *Fuchsia* and carrot, the presequence of *rps10* in *Arabidopsis*⁹ has no similarity to any sequences currently in BLAST databases at the National Center for Biotechnology Information, which suggests that it was an independent acquisition.

In contrast to gene activation mediated by an acquired presequence, nuclear *rps10* genes from spinach, maize and *Oxalis* do not contain upstream extensions of the open reading frame. When incubated with isolated mitochondria, each of these RPS10 proteins was imported to a protease-insensitive location in a membrane-potential-dependent manner (Fig. 4), strongly suggesting import into the mitochondrion despite the lack of a presequence. Nuclear *rps10* in rice (clearly derived from the same transfer as maize *rps10*) lacks a presequence and its product has been shown to be present in rice mitochondria in a concurrent study¹⁰, further supporting the inference that maize RPS10 proteins are present in mitochondria *in vivo*. Import without a presequence is rare in plants^{11–13}, and unprecedented following recent gene transfer. Targeting information must be present in the mitochondrial *rps10*-homologous region in these plants. Thus, *rps10* was apparently activated in spinach, maize and *Oxalis* without gaining a presequence, bypassing one of the normal requirements for activation. Notably, maize *rps10* contains an intron in the 5' untranslated region (UTR) (as does *rps10* in rice¹⁰) suggesting the acquisition of 5' regulatory elements by an exon-shuffling-like process. Nuclear *rps10* from lettuce is a curious blend of a gene fusion without a cleavable presequence. Lettuce *rps10* is present as an in-frame insertion within a gene for a non-mitochondrial metalloprotease (Fig. 3c), and lettuce RPS10 was imported into isolated mitochondria without cleavage of its

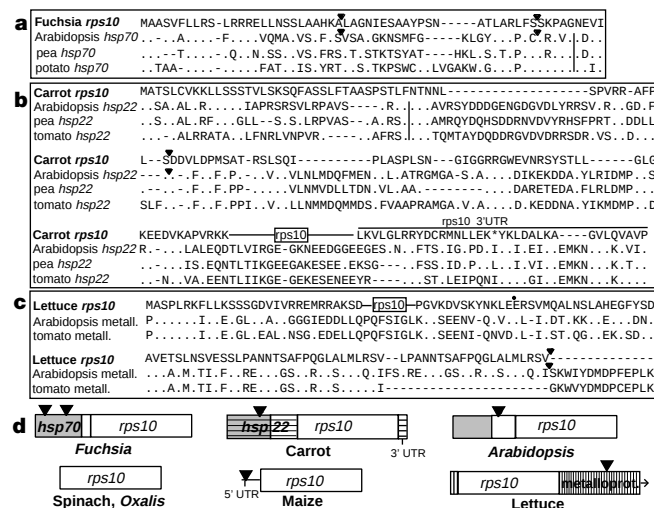


Figure 3 Partial sequence comparisons of three chimaeric nuclear *rps10* genes and structures of seven nuclear *rps10* genes. Triangles mark intron positions. **a–c**, Sequences from *Arabidopsis* are genomic sequences; all other non-*rps10* sequences are cDNAs (see Supplementary Information). Dots indicate amino acids identical to the top (*rps10*) sequence; dashes indicate gaps inserted to improve alignment. Vertical lines indicate the cleavage site of HSP70 and HSP22 presequences. The RPS10-homologous region is indicated by a box at its insertion site. The carrot *rps10* 3' UTR has been shown as a translation for alignment purposes only. The bullet in **c** indicates the amino acid in lettuce RPS10 that was changed to a stop codon for import experiments. **d**, Structures of seven nuclear *rps10* genes. Mitochondrial targeting sequences are shaded; horizontal or vertical stripes indicate regions homologous to other nuclear genes. Arrow indicates continuing open reading frame. Short N-terminal extensions of the RPS10 mature coding region are present in *Fuchsia*, carrot, *Arabidopsis* and lettuce.

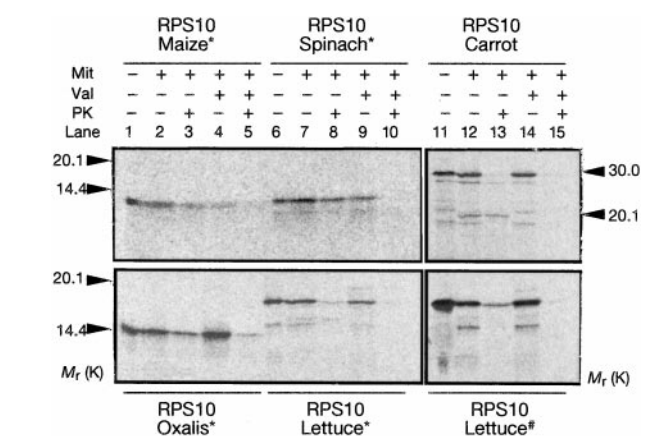


Figure 4 Import of RPS10 proteins into isolated mitochondria. Lane 1, ³⁵S-labelled precursor protein; lane 2, precursor incubated with isolated soybean mitochondria; lane 3, as in lane 2 except that proteinase K was added after mitochondrial incubation to degrade unimported proteins; lanes 4 and 5, as in lanes 2 and 3, respectively, except that valinomycin was added before import to dissipate the membrane potential; lanes 6–10 and 11–15 are equivalent to 1–5. Markers are given in relative molecular mass (*M_r* (K)). Asterisks indicate those proteins that required denaturation before import. Hash indicates incubation with potato tuber mitochondria instead of soybean mitochondria. As a negative control, a chloroplast protein (rubisco small subunit) did not import into either mitochondrial preparation (data not shown).

short N-terminal extension (Fig. 4). Finally, the reading frame of *Passiflora rps10* is fused at its 5' end with part of a gene of unknown function (data not shown); we have not yet determined whether the 5' extension is a presequence or whether the chimaeric structure extends to the 3' end of *rps10*.

The disparate structures of nuclear *rps10* in *Fuchsia*, carrot, *Arabidopsis*, lettuce and perhaps *Passiflora* imply that each gene was derived from a separate transfer event, with the various gene fusions, insertions and presequence acquisitions probably occurring at the same time as or soon after transfer to the nucleus. The distinctive gene structure of *Oxalis rps10* relative to that of the other three rosids (*Fuchsia*, *Arabidopsis* and *Passiflora*), along with its low silent and non-silent substitution values relative to those of other nuclear *rps10* genes (Supplementary Information; see also Fig. 5a) suggests that *Oxalis rps10* was also derived from a separate, and recent, transfer. The sporadic distribution of the 26 losses of mitochondrial *rps10*, combined with the presence of intact mitochondrial *rps10* genes in species that are very closely related, imply that the losses have occurred relatively recently. Mitochondrial gene loss has been shown to occur relatively soon after transfer to the nucleus³; thus, recent losses imply recent transfers. On the basis of the above reasoning, there appear to have been at least seven separate transfers of *rps10* to the nucleus and probably many more; each of the 16 nuclear *rps10* genes may represent a separate case of functional transfer to the nucleus.

Phylogenetic analyses provided an additional source of evidence for evaluating the number and timing of *rps10* transfers. The utility of phylogenetic analyses was limited by the short length (~330 base pairs) of the *rps10* gene and the large difference in nucleotide substitution rate between mitochondrial and nuclear genes in plants^{4,5}. There are few shared derived changes uniting groups of mitochondrial *rps10* sequences, and these could easily have been eliminated after transfer to the nucleus owing to its high rate of nucleotide substitutions. Nonetheless, we performed maximum likelihood analyses using 32 mitochondrial *rps10* sequences, representing a diverse range of angiosperms, and 1–7 nuclear *rps10* sequences (Fig. 5; Supplementary Information and data not shown). Analyses that included all seven nuclear sequences produced two biologically uninterpretable clusters of long-branched nuclear sequences (Fig. 5a). To mitigate long-branch artefacts, nuclear *rps10* sequences were analysed singly. Four (*Fuchsia*, *Arabidopsis*, lettuce and *Oxalis*) branched with a mitochondrial

rps10 from a closely related species whereas three others did not (data not shown). When the nuclear *rps10* sequences from *Arabidopsis*, *Fuchsia* and lettuce were analysed together, they branched at the same positions as in the single analyses (Fig. 5b). These results support the hypothesis that the *Fuchsia*, lettuce and *Arabidopsis* nuclear *rps10* genes are the result of three separate, and recent, transfers from the mitochondrion.

Although the available data point to a separate transfer event for each of the seven well-characterized nuclear *rps10* genes, alternative hypotheses cannot be ruled out. One alternative is a single transfer in a common ancestor of monocots and eudicots, with the transferred gene's product initially targeted to mitochondria without a presequence (as in maize, spinach and *Oxalis*). The unique presequences and fusions with other genes would be explained as recent secondary derivations through genomic rearrangement or re-insertion of complementary DNAs. This seems unlikely to have occurred on the scale found here (4 out of 7 cases), because gain of a new presequence after ancient transfer is rare¹⁴ and because retention of presumably functionally redundant copies of *rps10* in the nucleus and mitochondrion during much of angiosperm evolution seems improbable^{3,15}. The recency of the 26 losses of mitochondrial *rps10* and the phylogenetic results are also more difficult to explain by one or a few ancient transfers.

We have found evidence for a high rate of *rps10* loss from the mitochondrion and transfer to the nucleus during recent angiosperm evolution. It is possible that each of the 26 cases of *rps10* loss represents a separate transfer to the nucleus. Our study includes about 2% of extant angiosperm genera; therefore, *rps10* may have been independently transferred hundreds of times during angiosperm evolution. Is *rps10* exceptional, perhaps because it may not always need a presequence to be activated in the nucleus, or have other mitochondrial genes also been transferred many times in angiosperms? Hybridization surveys reveal numerous losses of many other mitochondrial ribosomal protein genes (our own unpublished data); whether these also reflect numerous transfers is not yet known, although two separate transfers of *rps14* have been inferred^{7,16}. In contrast to plants, only a single protein gene (*atp8*) has been lost (twice) among over 100 sequenced mitochondrial genomes from animals², and these two losses may not even reflect transfer to the nucleus (there is no detectable *atp8* sequence in the nucleus of *Caenorhabditis elegans*). Non-functional transfer/duplication of mitochondrial sequences to the nucleus is, however,

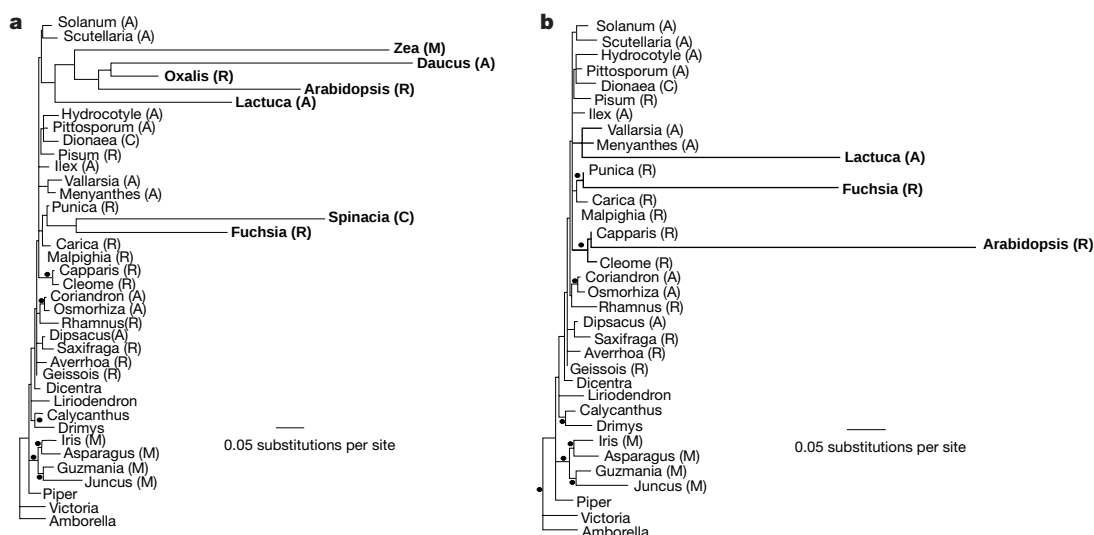


Figure 5 Maximum likelihood trees of mitochondrial and nuclear *rps10* nucleotide sequences. **a**, Thirty-two mitochondrial and seven nuclear (in bold) *rps10* sequences. **b**, Thirty-two mitochondrial and three nuclear *rps10* nucleotide sequences. Letters in

parentheses indicate major phylogenetic groups: R, rosids; A, asterids; C, caryophyllids; and M, monocots. Bullets indicate all bootstrap values above 40%. See Supplementary Information for alternative topology testing.

common in animals (for example, see ref. 17); the lack of functional transfer may reflect a stringent barrier imposed by genetic code incompatibilities in animals but not plants. Multiple, probably mostly ancient, losses have been inferred for many mitochondrial and chloroplast genes across the broad sweep of eukaryotic evolution^{18–20}. However, corresponding nuclear genes have generally been isolated from at most a single loss lineage, and therefore the incidence of (parallel) gene transfer is unclear.

The high rates of *rps10* loss and transfer in angiosperms seem to equal or exceed rates of two much simpler and, one would think, much more likely classes of mitochondrial mutations in angiosperms. These include intron loss (Fig. 1; our own unpublished data) and perhaps even the trivial mutation of substitution at a silent site in a protein gene. This is surprising, considering that functional gene transfer is such a complex, multistep process, involving reverse transcription of a mitochondrial messenger RNA, movement to the nucleus, chromosomal integration, gain of a nuclear promoter and other regulatory elements, gain of a presequence (usually), and, ultimately, mitochondrial gene loss. Relocation of *rps10* to the nucleus is occurring at a dizzying pace in angiosperms, with the many cases showing a remarkable range of opportunistic gene fusions and co-options leading to functional activation with or without gain of a mitochondrial presequence. □

Methods

Total cellular DNA and RNA were isolated as described³. For Southern blot hybridization, total genomic DNAs were cut with *HindIII*, electrophoresed, blotted and hybridized³ at moderate stringency (60 °C in 5 × SSC; washes at 60 °C in 2 × SSC). We isolated nuclear (16) and mitochondrial (30) *rps10* genes by PCR, reverse transcription-PCR and 5' rapid amplification of cDNA ends as described³ (see Supplementary Information for primers and inverse PCR). All PCR products were sequenced directly, or cloned using the TA cloning kit (Invitrogen) followed by sequencing of multiple clones. Maximum likelihood analyses were conducted using PAUP*4.0b3a (ref. 21). We used the HKY85 model, assuming a discrete gamma distribution with four categories of site-to-site rate variability. For each analysis, we estimated the transition/transversion ratio and base frequencies using Tree-Puzzle (version 4.02 (ref. 22)) under the HKY model of evolution, with gamma-distributed rates and parameter estimation set to 'approximate'. We excluded the five RNA editing sites. Our analyses used heuristic searches with random-taxon addition (10 replicates) and TBR branch-swapping. We carried out all analyses at least twice until a stable topology was achieved. Bootstrapping was done using PAUP* as above, with stepwise addition and 100 replicates.

The mitochondria from soybean cotyledons²³ and potato tubers²⁴ were prepared. ³⁵S-labelled RPS10 precursor proteins were synthesized from cDNA clones and imported into the isolated mitochondria *in vitro*²⁵. In some cases, smaller precursor proteins were denatured before import²⁶.

Received 11 May; accepted 16 August 2000.

- Gray, M. W. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* **141**, 233–357 (1992).
- Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
- Adams, K. L. *et al.* Intracellular gene transfer in action: Dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. *Proc. Natl Acad. Sci. USA* **96**, 13863–13868 (1999).
- Wolfe, K. H., Li, W.-H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl Acad. Sci. USA* **84**, 9054–9058 (1987).
- Laroche, J., Li, P., Maggia, L. & Bousquet, J. Molecular evolution of angiosperm mitochondrial exons and introns. *Proc. Natl Acad. Sci. USA* **94**, 5722–5727 (1997).
- Kadowaki, K., Kubo, N., Ozawa, K. & Hirai, A. Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. *EMBO J.* **15**, 6652–6661 (1996).
- Figuerola, P., Gómez, L., Holuigue, L., Araya, A. & Jordana, X. Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J.* **18**, 601–609 (1999).
- Kubo, N., Harada, K., Hirai, A. & Kadowaki, K. A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: Common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl Acad. Sci. USA* **96**, 9207–9211 (1999).
- Wischmann, C. & Schuster, W. Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. *FEBS Lett.* **375**, 152–156 (1995).
- Kubo, N. *et al.* Transfer of the mitochondrial *rps10* gene to the nucleus in rice: acquisition of the 5' untranslated region followed by gene duplication. *Mol. Gen. Genet.* **263**, 733–739 (2000).
- Nakagawa, T., Maeshima, M., Nakamura, K. & Asahi, T. Molecular cloning of a cDNA for the smallest nuclear-encoded subunit of sweet potato cytochrome *c* oxidase. *Eur. J. Biochem.* **191**, 557–561 (1990).
- Morikami, A., Ehara, G. & Yuuki, K. Molecular cloning and characterization of cDNAs for the γ -subunit and ϵ -subunit of mitochondrial F1F0 ATP synthase from sweet potato. *J. Biol. Chem.* **268**, 17205–17210 (1993).
- Braun, H.-P., Jansch, L., Kruff, V. & Schmitz, U. K. The 'Hinge' protein of cytochrome *c* reductase from potato lacks the acidic domain and has no cleavable presequence. *FEBS Lett.* **347**, 90–94 (1994).

- Long, M., de Souza, S. J., Rosenberg, C. & Gilbert, W. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome *c1* precursor. *Proc. Natl Acad. Sci. USA* **93**, 7727–7731 (1996).
- Herrmann, R. G. in *Eukaryotism and Symbiosis* (eds Schenk, H. E. A. *et al.*) 73–118 (Springer, Vienna, 1997).
- Figuerola, P. *et al.* The gene for mitochondrial ribosomal protein S14 has been transferred to the nucleus in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **262**, 139–144 (1999).
- Bensasson, D., Zhang, D.-X. & Hewitt, G. M. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol. Biol. Evol.* **17**, 406–415 (2000).
- Palmer, J. D. The mitochondrion that time forgot. *Science* **275**, 790–791 (1997).
- Martin, W. *et al.* Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165 (1998).
- Gray, M. W. Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* **9**, 678–687 (1999).
- Swofford, D. L. PAUP*, *Phylogenetic Analysis Using Parsimony (*and Other Methods)* 4.0B3 edn (Sinauer Associates, Sunderland, MA, 2000).
- Strimmer, K. & von Haeseler, A. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969 (1996).
- Day, D. A., Neuburger, M. & Douce, R. Biochemical characterization of chlorophyll-free mitochondria from pea leaves. *Aust. J. Plant Phys.* **12**, 219–228 (1985).
- Neuburger, M., Journet, E., Bligny, R., Carde, J. & Douce, R. Purification of plant mitochondria by isopycnic centrifugation in density gradients of percoll. *Arch. Biochem. Biophys.* **217**, 312–323 (1982).
- Tanudji, M., Sjoling, S., Glaser, E. & Whelan, J. Signals required for the import and processing of the alternative oxidase into mitochondria. *J. Biol. Chem.* **274**, 1286–1293 (1999).
- Knox, C., Sass, E., Neupert, W. & Pines, O. Import into mitochondria, folding, and retrograde movement of fumarate in yeast. *J. Biol. Chem.* **273**, 25587–25593 (1998).
- Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**, 402–404 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank W. Fischer, J. Logsdon, C. Parkinson, M. Rosenbluth, N. Schisler and K. Wolfe for reading the manuscript, and D. Swofford for allowing us to use a pre-release version of PAUP*. This study was supported by a United States Department of Agriculture graduate fellowship to K.L.A., grants to Y.L.Q. and J.D.P. from the NIH, and grants from the Australian Research Council to J.W.

Correspondence and requests for materials should be addressed to J.D.P. (e-mail: jpalm@bio.indiana.edu). The sequences reported in this study have been deposited in GenBank under the accession numbers AF287307–AF287345.

Imagery neurons in the human brain

Gabriel Kreiman*, Christof Koch* & Itzhak Fried†

* Computation and Neural Systems Program, California Institute of Technology, 139-74, Pasadena, California 91125, USA

† Division of Neurosurgery, and Department of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles, School of Medicine, 740 Westwood Plaza, Los Angeles, California 90095, USA

Vivid visual images can be voluntarily generated in our minds in the absence of simultaneous visual input. While trying to count the number of flowers in Van Gogh's *Sunflowers*, understanding a description or recalling a path, subjects report forming an image in their "mind's eye"¹. Whether this process is accomplished by the same neuronal mechanisms as visual perception has long been a matter of debate^{1–3}. Evidence from functional imaging^{1,4–8}, psychophysics^{1,9}, neurological studies² and monkey electrophysiology^{10–12} suggests a common process, yet there are patients with deficits in one but not the other^{3,13}. Here we directly investigated the neuronal substrates of visual recall by recording from single neurons in the human medial temporal lobe^{14,15} while the subjects were asked to imagine previously viewed images. We found single neurons in the hippocampus, amygdala, entorhinal cortex and parahippocampal gyrus that selectively altered their firing rates depending on the stimulus the subjects were imagining. Of the neurons that fired selectively during both vision and imagery, the majority (88%) had identical selectivity. Our study reveals single neuron correlates of volitional visual imagery in humans and

suggests a common substrate for the processing of incoming visual information and visual recall.

We directly studied the neuronal correlates of visual imagery by recording from single neurons in the human brain. Subjects were nine patients with pharmacologically intractable epilepsy implanted with chronic electrodes to localize the seizure foci for possible surgical resection¹⁴. Based on clinical criteria, electrodes were implanted bilaterally in the amygdala, entorhinal cortex, hippocampus and parahippocampal gyrus. We compared the selectivity of the neurons during vision and visual imagery for different stimuli. Two images were separately shown for 1,000 ms per presentation and five repetitions per image (Fig. 1a, b). Figures were drawn from the following nine groups: faces showing emotions, household objects, spatial layouts, cars, animals, drawings and photographs of famous people, foodstuffs and complex patterns¹⁵. Subsequently, the subjects closed their eyes and imagined one of the two pictures shown upon listening to a high or low tone (Fig. 1d). Tones were alternated every 3,000 ms and there were five repetitions of the tones per image. Visual imagery was verified by debriefing after each pair of pictures.

We recorded from 276 single neurons in the medial temporal lobe (Table 1). We found that some of the neurons showed selective changes in firing rate while subjects viewed the figures and when they were visually recalling the images with closed eyes. A neuron was considered to be selective to one of the stimulus groups if: (1) the firing rate during stimulus presentation was significantly different from the baseline activity (Wilcoxon test); (2) the response

was different from that to stimuli from all other stimulus groups (analysis of variance (ANOVA) and pairwise Wilcoxon comparisons); and (3) no significant differences were found in the response to distinct individual stimuli within the group (ANOVA).

The baseline activity of the neurons during vision was computed in the 1,000-ms interval before each presentation. During visual imagery, we avoided comparing to a baseline between tones when subjects could still be imagining the stimuli; instead, we used the 1,000-ms interval before the first tone. Neuronal activity during the baseline constitutes a potential concern in visual imagery experiments^{1,5,7}. There was no significant difference in the firing rate of any of the neurons in the vision and imagery baselines (Wilcoxon test, $P > 0.2$). Furthermore, both baselines were indistinguishable from the spontaneous activity of the neurons computed over the entire experimental session.

Figure 2a shows an example of activity recorded from a neuron in the entorhinal cortex that increased its firing rate selectively when the subject viewed pictures of objects. The mean firing rate during the interval between 100 and 1,000 ms after stimulus onset for objects was 16.8 ± 3.6 (mean $\pm$ s.d.) spikes per second. This was significantly higher than the baseline and also higher than the activity for all other types of stimulus (ANOVA and pairwise comparisons, $P < 10^{-3}$). The same neuron also increased its firing rate when the patient recalled the objects with closed eyes (Fig. 2b), but not when the subject imagined other stimuli. Activity recorded from another example of a neuron that showed selective changes in firing rate during vision and visual recall is shown in Fig. 3. This

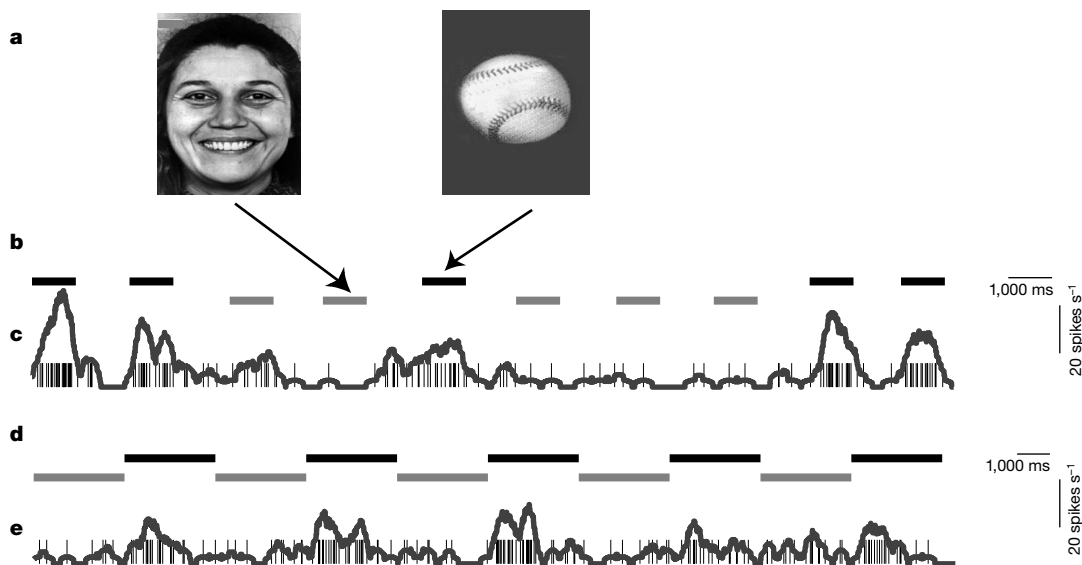


Figure 1 Individual responses of a single neuron during vision and imagery. **a**, Two images were shown separately for 1,000 ms each, five repetitions per image, indicated by horizontal black and white bars in **b**. After each picture, subjects pressed a button indicating whether the picture was a human face. **c**, Activity from a neuron in the entorhinal cortex; the continuous line shows the spike density function. After 10 visual

presentations, subjects closed their eyes and imagined one or the other picture upon hearing a high or low tone. **d**, Tones were alternated every 3,000 ms. **e**, Data from the same neuron during visual imagery. This neuron showed an increased firing rate for pictures of objects ($P < 10^{-3}$) during both vision and imagery.

Table 1 Number of responsive and selective neurons

	Amygdala	Entorhinal cortex	Hippocampus	Parahippocampal gyrus	Total
<i>n</i>	89	78	91	18	276
Visual responsive	12 (13%)	16 (21%)	17 (19%)	4 (22%)	49 (18%)
Visual selective	9 (10%)	14 (18%)	17 (19%)	4 (22%)	44 (16%)
Imagery responsive	8 (9%)	11 (14%)	9 (10%)	5 (28%)	33 (12%)
Imagery selective	4 (4%)	8 (10%)	8 (9%)	3 (17%)	23 (8%)
Both selective	3 (75%)	6 (75%)	5 (63%)	2 (67%)	16 (70%)

Number of responsive and selective neurons detected out of the number of neurons recorded in each location (*n*). The percentages for the responsive and selective neurons are based on the total number of recorded neurons. 'Both selective' indicates those neurons selective during vision and imagery. For these, the percentages are based on the total number of imagery selective neurons. Of the 16 neurons selective during both processes, 14 showed the same selectivity. The neurons were selective for faces, objects, spatial layouts and other stimuli. A χ^2 test¹⁴ to address the probability of obtaining the number of selective neurons by chance yielded $P < 0.01$ for the 28 vision-only neurons, $P = 0.04$ for the 7 imagery-only neurons and $P < 10^{-6}$ for the 16 neurons selective during both.

amygdala neuron (from a different patient) showed an increased firing rate when the subject saw pictures of animals and when she formed mental images of the same pictures, but not during vision or recall of other stimuli.

We found a total of 44 neurons (16% of all recorded neurons) that showed selective changes in firing rate during visual presentation (Table 1). There were 23 neurons (8%) that showed selective changes in firing rate while subjects visually imagined the same stimuli. Of these 23 neurons, 7 (30%) were activated exclusively during imagery whereas 16 (70%) were selective during vision. Remarkably, of these 16 neurons, 14 (88%) showed the same selectivity during vision and visual imagery. While the significance criterion was set to 0.05, most of the *P* values (vision: 78%; imagery: 67%) were below 0.01. Assuming a null hypothesis of independence between vision and imagery, at the 0.01 level we would expect approximately one neuron in 90,000 to show the same selectivity during both vision and imagery by chance. We found 10 neurons in

our sample of 276 neurons with $P < 0.01$ (and 14 with $P < 0.05$) that responded with the same selectivity during vision and imagery. As is consistent with the longer period of the imagery task and the more temporally diffuse nature of imagery, the latencies and the time of peak activity were longer and more variable for imagery than for vision (latency was 282 ± 191 ms for vision and 409 ± 291 ms for imagery; peak time was 665 ± 247 ms for vision and $1,482 \pm 921$ ms for imagery).

Does the firing rate of the neurons during vision differ during imagery of the same stimuli? To address this question, for the 14 neurons with the same selectivity we computed the firing rate over the whole stimulus period for the selective stimuli and divided it by the baseline activity. There was a strong correlation between this normalized firing response for vision and visual imagery ($r^2 = 0.90$). The slope between the activity during vision and imagery was 0.74, indicating that the firing rates were about 25% higher during vision. As the duration of the selective response was typically shorter than

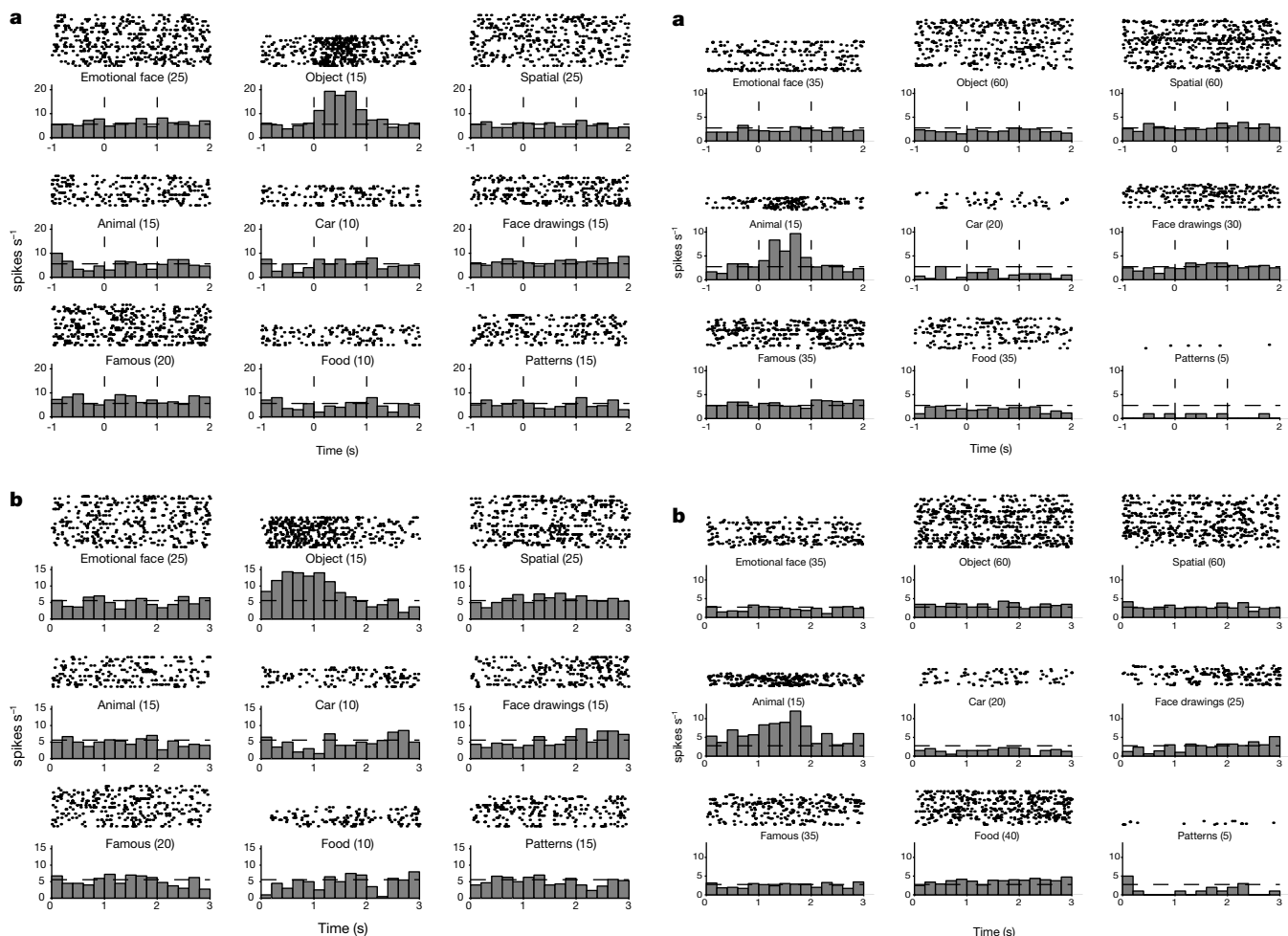


Figure 2 Responses of the same neuron as in Fig. 1 during vision and visual imagery. **a**, During vision; **b**, during visual imagery. The post-stimulus time histograms were computed by averaging activity for all stimuli within each stimulus group (the total number of presentations is indicated in parentheses) using a bin size of 200 ms. This neuron increased its firing rate over the baseline ($P < 10^{-4}$) upon visually presenting objects but not other stimuli. An ANOVA and pairwise comparisons indicated that the response to objects was significantly different from that to other stimuli ($P < 10^{-3}$). The neuron also increased its activity when the subject recalled the same objects in her mind with eyes closed (comparison with baseline, ANOVA and pairwise comparisons: $P < 0.001$) but not during imagery of other stimuli. There was no significant difference in the responses to distinct objects (vision $P > 0.2$; imagery $P > 0.2$).

Figure 3 Responses of a selective neuron in the left amygdala of a different subject during vision and visual imagery. **a**, During vision; **b**, during visual imagery. This neuron increased its firing rate over the baseline upon visually presenting animals ($P < 10^{-5}$) and also during imagery of animals ($P < 10^{-4}$) but not to other stimuli. ANOVA and pairwise comparisons also showed that the response of this neuron was highly selective during both vision ($P < 10^{-3}$) and visual imagery ($P < 10^{-3}$). There was no significant difference in the responses to distinct animals (vision $P > 0.15$; imagery $P > 0.3$).

the whole stimulus period (mean 560 ± 292 ms for vision and 948 ± 580 ms for imagery), this firing rate is an underestimation of the response of the neuron. Furthermore, the intervals in the two tasks were different. We therefore computed the firing rates in a 600-ms period centred on the peak firing rate. The correlation coefficient between the firing rate in this interval during vision and imagery was 0.95 and the slope was 0.85 (Fig. 4a). Given the weaker percept during imagery, it may seem surprising that the difference in firing responses is so small. Yet, in our sample there were fewer cells recruited during visual imagery than during vision (Table 1). Similar results were observed by functional imaging studies⁷. Selectivity to one of the categories may be due to physical similarity between the stimuli. The correlation of firing rates between vision and imagery was also high for individual figures ($r^2 = 0.93$, slope is 0.88; see for example Fig. 1.) Therefore, the correlation does not rely on category selectivity. Furthermore, there was no significant difference between the responses to distinct individual stimuli within the selective group for the selective neurons (ANOVA, vision $P > 0.1$; imagery $P > 0.15$).

Is it possible for an ideal observer to predict the group of the stimulus that the subject was viewing or imagining on the basis of the firing rate of a single neuron? Figure 1 shows a particularly clear example of selective firing in a single repetition of the neuron whose average activity was depicted in Fig. 2. By observing the activity of this neuron, it is possible to predict with rather high accuracy what the subject was viewing (Fig. 1c) or imagining (Fig. 1e). We addressed this question quantitatively by carrying out a receiver operating characteristic analysis¹⁶. This yielded the minimum probability of error, P_e , in classifying the stimulus as belonging to the preferred category or not on the basis of the average firing rate¹⁵. $P_e = 0$ corresponds to perfect classification and $P_e = 0.5$ to chance performance. P_e ranged from 0.10 to 0.44 for vision (0.28 ± 0.07) and from 0.08 to 0.46 for visual imagery (0.27 ± 0.06). There was a strong correlation in the P_e values for the neurons selective in both vision and imagery (Fig. 4b).

Extrastriate areas in the human brain are specialized for processing complex visual input. For instance, functional magnetic resonance imaging studies show areas specialized for faces¹⁷ and places¹⁸. Neurons in monkey extrastriate cortex respond selectively to complex stimuli^{19–22}. Their activity represents the pictorial short-term memory^{11,23,24} and neuronal correlates of visual recall and

prospective coding have been described in monkeys^{10–12}. Furthermore, these studies show top-down interactions between prefrontal and temporal cortex during recall. These neurons project to the medial temporal lobe^{20,25} where lesions lead to specific visual deficits in both macaques and humans^{2,17,19,20,26,27}. Furthermore, electrical stimulation in the human temporal lobe can interfere with^{28,29} or elicit visual recall³⁰.

We observed three different types of selective neurons. Some neurons responded during processing of incoming visual information but not during imagery (28/44). There were also neurons activated only during visual recall (7/23), which may be involved in retrieval mechanisms dissociated from vision^{3,13}. Finally, some neurons responded selectively during both vision and imagery (16/23). Of these 16 neurons, 14 showed identical selectivity. There are patients with deficits in both vision and imagery^{1,2}, but some neurological lesions yield impairments in one but not the other^{3,13}. There is considerable convergence of input to the medial temporal lobe and it is plausible that these lesions involve neuronal systems that project to neurons in the areas studied here. We did not find regional segregation of neurons selective during vision and imagery; both were found in all regions (Table 1). Although our data suggest shared neuronal representation in the medial temporal lobe, it does not rule out the possibility of segregation between these processes.

Our results provide a rare opportunity to observe the activity of single neurons in the human brain directly in the absence of external visual stimulation. This activity may represent the retrieval of the picture information from memory or the maintenance of the visual percept during imagination. The firing of these neurons could represent a correlate of the percept common to vision and imagery. Given the prominent role of the medial temporal lobe in declarative memory, it also seems possible that these neurons could be activated during storage of incoming visual inputs and later reactivated during the mnemonic retrieval process required for imagery. □

Methods

Subjects and electrode implantation

Subjects were nine patients (21–42 years old, four male, seven right-handed) with pharmacologically intractable epilepsy. Extensive non-invasive monitoring did not yield concordant data corresponding to a single resectable focus and, therefore, they were implanted with chronic depth electrodes for 1–2 weeks to determine the seizure focus for possible surgical resection^{14,15}. The surgeries were performed by I.F. All studies conformed with the guidelines of the Medical Institutional Review Board at UCLA. Four of the patients also participated in a previous study where only visual responses were examined¹⁵. The location of the electrodes was verified by structural magnetic resonance imaging. The electrode locations were based exclusively on clinical criteria. The recordings during seizures were used to localize the focus. We note that generalization about normal neuronal function from recordings in epileptic patients constitutes a potential limitation. However, 87% of the recorded neurons were outside the clinically determined epileptogenic zone and we did not observe differences in waveforms or response properties in neurons near the seizure focus.

Tasks and recordings

During vision, two images were separately shown for 1,000 ms (Fig. 1a and b). In the first two patients, only faces, objects and spatial layouts were presented. After each picture, a tone reminded the subjects to press a button indicating whether the picture was a human face. Subsequently, subjects closed their eyes and imagined one or the other picture upon listening to high and low tones alternated every 3,000 ms (Fig. 1d). This was repeated for approximately 30 different pairs of images during each session, 2–4 sessions per patient, depending on clinical constraints. Imagery was verified by debriefing following each repetition by requesting detailed descriptions and asking whether subjects could form visual images. Each image appeared in only one pair. Three neurons (none selective) showed a different activity to high and low tones per se. We compared the activity before and after the behavioural response and the peri-response activity to the baseline. None of the visual or imagery neurons showed firing related to pressing the button. We also observed selectivity previously without behavioural responses¹⁴.

Data from each recorded microwire were amplified, high-pass filtered and stored for off-line cluster separation (Datawave). Because the microelectrodes were chronically implanted, no selection of neurons by moving the electrodes was performed. Typically, neurons recorded from the same microwire as a selective neuron were not selective. We did not find evidence of spatial segregation of selective responses within any region. Because the location of the electrodes was fixed and based on clinical criteria, we were not able to address what happens in lower visual areas.

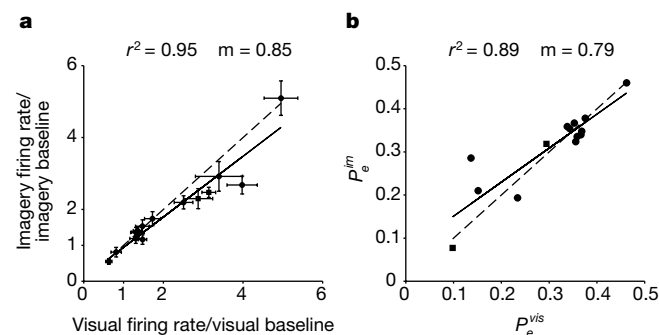


Figure 4 Comparison of firing rates and P_e between vision and imagery. **a**, Correlation of firing rates between vision and visual imagery for the 14 neurons that showed the same selectivity (Table 1). The firing rate was computed in a 600-ms window centred on the response peak and normalized by the baseline. The dashed line indicates $y = x$. The solid line shows a linear fit (slope is 0.85; correlation coefficient is 0.95; r^2 between the absolute firing rates is 0.93 and r^2 between the firing rates after subtracting the baseline is 0.97). The error bars correspond to standard deviation. **b**, Probability of error (P_e) in predicting the visual or the imagined percept on the basis of the firing rate of a single neuron. The P_e ranges from 0 (perfect classification) to 0.5 (chance performance). The dashed line shows $y = x$. The solid line indicates the linear fit (slope is 0.79; $r^2 = 0.89$). The examples from the previous figures are indicated by squares.

Data analysis

A neuron was considered selective to a stimulus group if: (1) the firing rate during stimulus presentation was different from the preceding baseline (Wilcoxon test, < 0.05), (2) an analysis of variance and pairwise comparisons (Wilcoxon test) addressing whether there were differences among the stimulus groups yielded $P < 0.05$ and (3) an ANOVA (parametric and non-parametric) comparing the variability to distinct stimuli within the selective category to the variability to repeated presentations of the same stimulus showed $P > 0.05$. We observed neurons selective to faces, objects, spatial layouts and other stimuli.

If the across-groups comparisons were not significant but the activity was different from baseline, the neuron was defined as responsive but non-selective. To take into account any effects due to the different intervals we also compared the responses in a 600-ms window centred on the peak firing rate. The peak, latency and duration were estimated from the spike density function¹⁵. For the selective neurons we computed the probability of error, P_e , for classifying the stimulus as belonging to the preferred stimulus category or not^{15,16}. We did not observe any difference between the right and left hemisphere neurons.

Received 21 July; accepted 22 August 2000.

- Kosslyn, S. M. *Image and Brain* (MIT Press, Cambridge, 1994).
- Farah, M. J. Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychol. Rev.* **95**, 307–317 (1988).
- Behrmann, M., Winocur, G. & Moscovitch, M. Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature* **359**, 636–637 (1992).
- Kosslyn, S. M., Thompson, W. L. & Alpert, N. M. Neural systems shared by visual imagery and visual perception: a PET study. *Neuroimage* **6**, 320–334 (1997).
- Roland, P. E. & Gulyas, B. Visual imagery and visual representation. *Trends Neurosci.* **17**, 281–287 (1994).
- D'Esposito, M. *et al.* A fMRI study of mental image generation. *Neuropsychologia* **35**, 725–730 (1997).
- O'Craven, K. & Kanwisher, N. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cog. Neurosci.* (in the press).
- Frith, C. & Dolan, R. J. Brain mechanisms associated with top-down processes in perception. *Phil. Trans. R. Soc. Lond.* **352**, 1221–1230 (1997).
- Ishai, A. & Sagi, D. Common mechanisms of visual imagery and perception. *Science* **268**, 1772–1774 (1995).
- Rainer, G., Rao, S. & Miller, E. Prospective coding for objects in primate prefrontal cortex. *J. Neurosci.* **19**, 5493–5505 (1999).
- Miyashita, Y. & Chang, H. S. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* **331**, 68–71 (1988).
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I. & Miyashita, Y. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* **401**, 699–703 (1999).
- Bartolomeo, P. *et al.* Multiple-domain dissociation between impaired visual perception and preserved mental imagery in a patient with bilateral extrastriate lesions. *Neuropsychologia* **36**, 239–249 (1998).
- Fried, I., MacDonald, K. A. & Wilson, C. Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* **18**, 753–765 (1997).
- Kreiman, G., Koch, C. & Fried, I. Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neurosci.* **3**, 946–953 (2000).
- Green, D. & Swets, J. *Signal detection theory and psychophysics* (Wiley, New York, 1966).
- Kanwisher, N. & Moscovitch, M. The cognitive neuroscience of face processing: An introduction. *Cogn. Neurosci.* **17**, 1–11 (2000).
- Epstein, R. & Kanwisher, N. A cortical representation of the local visual environment. *Nature* **392**, 598–601 (1998).
- Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* **19**, 577–621 (1996).
- Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* **19**, 109–139 (1996).
- Gross, C. G. How inferior temporal cortex became a visual area. *Cereb. Cortex* **5**, 455–469 (1994).
- Rolls, E. Neural organization of higher visual functions. *Curr. Opin. Neurobiol.* **1**, 274–278 (1991).
- Miyashita, Y. Inferior temporal cortex: Where visual perception meets memory. *Annu. Rev. Neurosci.* **16**, 245–263 (1993).
- Chelazzi, L., Duncan, J., Miller, E. K. & Desimone, R. Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophys.* **80**, 2918–2940 (1998).
- Suzuki, W. A. Neuroanatomy of the monkey entorhinal, perirhinal and parahippocampal cortices: Organization of cortical inputs and interconnections with amygdala and striatum. *Semin. Neurosci.* **8**, 3–12 (1996).
- Warrington, E. & McCarthy, R. Categories of knowledge—Further fractionations and an attempted integration. *Brain* **110**, 1273–1296 (1987).
- Meunier, M., Hadfield, W., Bachevalier, J. & Murray, E. Effects of rhinal cortex lesions combined with hippocampotomy on visual recognition memory in rhesus monkeys. *J. Neurophysiol.* **75**, 1190–1205 (1996).
- Fried, I., Mateer, C., Ojemann, G., Wohns, R. & Fedio, P. Organization of visuospatial functions in human cortex. *Brain* **105**, 349–371 (1982).
- Ojemann, G. & Mateer, C. Human language cortex: localization of memory, syntax, and sequential motor-phoneme identification systems. *Science* **205**, 1401–1403 (1979).
- Penfield, W. & Jasper, H. *Epilepsy And The Functional Anatomy Of The Human Brain* (Little, Brown & Co., Boston, 1954).

Acknowledgements

This work was supported by grants from NIH, the Centre for Consciousness Studies at the University of Arizona and the Keck Foundation. We thank M. Zirlinger for discussions, T. Fields, C. Wilson, E. Isham and E. Behnke for help with the recordings, F. Crick for comments and I. Wainwright for editorial assistance. We also thank all the patients who participated in these studies.

Correspondence and requests for materials should be addressed to I.F. (e-mail: ifried@mednet.ucla.edu).

Real-time prediction of hand trajectory by ensembles of cortical neurons in primates

Johan Wessberg*, Christopher R. Stambaugh*, Jerald D. Kralik*, Pamela D. Beck*, Mark Laubach*, John K. Chapin†, Jung Kim‡, S. James Biggs‡, Mandayam A. Srinivasan‡ & Miguel A. L. Nicolelis*§||

* Department of Neurobiology; § Department of Biomedical Engineering;

|| Department of Psychology-Experimental, Duke University, Durham, North Carolina 27710, USA

† Department of Physiology and Pharmacology, State University of New York Health Science Center, Brooklyn, New York 11203, USA

‡ Laboratory for Human and Machine Haptics, Department of Mechanical Engineering and Research Laboratory of Electronics, MIT, Cambridge, Massachusetts 02139, USA

Signals derived from the rat motor cortex can be used for controlling one-dimensional movements of a robot arm¹. It remains unknown, however, whether real-time processing of cortical signals can be employed to reproduce, in a robotic device, the kind of complex arm movements used by primates to reach objects in space. Here we recorded the simultaneous activity of large populations of neurons, distributed in the pre-motor, primary motor and posterior parietal cortical areas, as non-human primates performed two distinct motor tasks. Accurate real-time predictions of one- and three-dimensional arm movement trajectories were obtained by applying both linear and nonlinear algorithms to cortical neuronal ensemble activity recorded from each animal. In addition, cortically derived signals were successfully used for real-time control of robotic devices, both locally and through the Internet. These results suggest that long-term control of complex prosthetic robot arm movements can be achieved by simple real-time transformations of neuronal population signals derived from multiple cortical areas in primates.

Several interconnected cortical areas in the frontal and parietal lobes are involved in the selection of motor commands for producing reaching movements in primates^{2–8}. The involvement of these areas in many aspects of motor control has been documented extensively by serial single-neuron recordings of primate behaviour^{2,3,8,9}, and evidence for distributed representations of motor information has been found in most of these studies^{10–13}, but little is known about how these cortical areas collectively influence the generation of arm movements in real time. The advent of multi-site neural ensemble recordings in primates¹⁴ has allowed simultaneous monitoring of the activity of large populations of neurons, distributed across multiple cortical areas, as animals are trained in motor tasks¹⁵. Here we used this technique to investigate whether real-time transformations of signals generated by populations of single cortical neurons can be used to mimic in a robotic device the complex arm movements used by primates to reach for objects in space.

Microwire arrays were implanted in multiple cortical areas of two owl monkeys (*Aotus trivirgatus*)^{14–16}. In the first monkey, 96 microwires were implanted in the left dorsal premotor cortex (PMd, 16 wires), left primary motor cortex (MI, 16 wires)^{17,18}, left posterior parietal cortex (PP, 16 wires), right PMd and MI (32 wires), and right PP cortex (16 wires). In the second monkey, 32 microwires were implanted in the left PMd (16 wires) and in the left MI (16 wires). Recordings of cortical neural ensembles began 1–2 weeks after the implantation surgery and continued for 12 months in monkey 1, and 24 months in monkey 2. During this period, the monkeys were trained in two distinct motor tasks. In task 1, animals

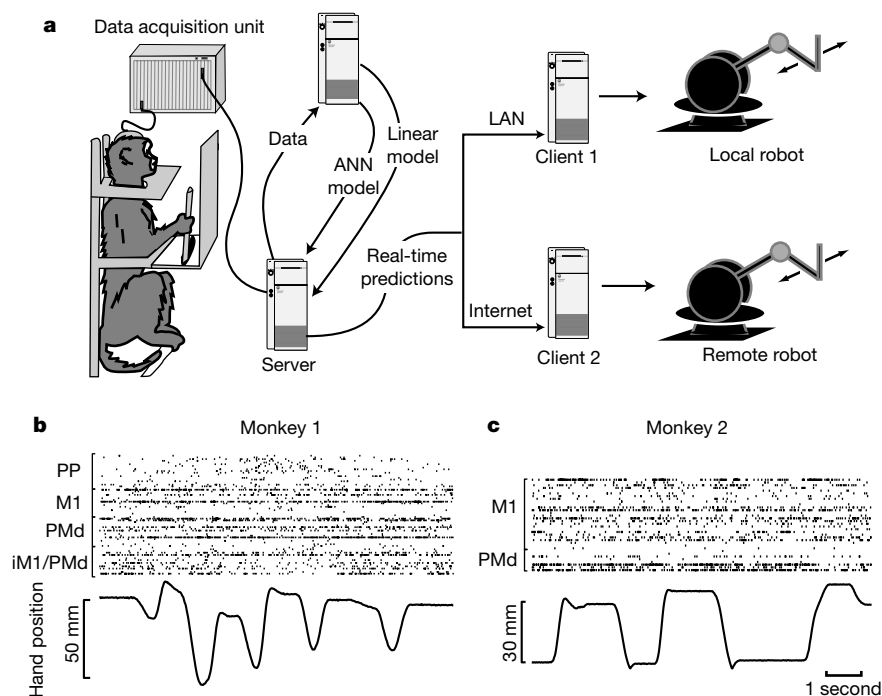


Figure 1 Experimental design. **a**, Schematic diagram depicting the experimental apparatus employed to use simultaneously recorded cortical ensemble data from primate behaviour to control the movements of local and remote robotic devices. In this design, both linear and ANN models are continuously updated during the recording session. Control of a local device is achieved through a local area network (LAN), while remote

control requires transmission of brain-derived signals through the Internet. **b, c**, Simultaneously recorded neuronal activity in five cortical areas in monkey 1 (**b**) and two cortical areas in monkey 2 (**c**) during the execution of 1D movements. PP, posterior parietal cortex; M1, primary motor cortex; PMd, dorsal premotor cortex; iM1/PMd, ipsilateral M1 and PMd.

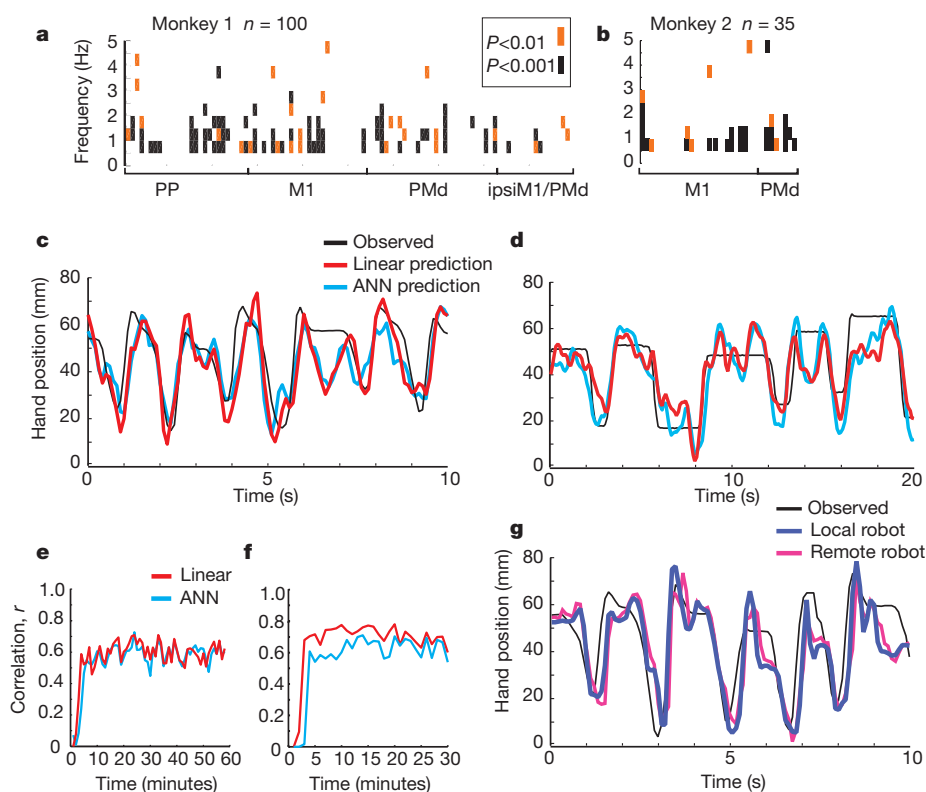


Figure 2 Real-time control of 1D hand movements. **a, b**, Coherence analysis reveals that most cortical neurons are significantly coupled with different frequency components of the movements. **c, d**, Observed (black) and real-time predicted 1D hand movements using linear (red) and ANN (blue) models in monkey 1 (**c**) and 2 (**d**). **e, f**, Correlation coefficient

variation for predicted hand movements, using linear (red) and ANN (black) models, in one recording session in monkey 1 (**e**) and 2 (**f**). **g**, Real-time 1D movements of a local (blue) and remote (red) robot arm obtained in monkey 1 by using the linear model.

made one-dimensional hand movements to displace a manipulator in one of two directions (left or right) following a visual cue. In task 2, the monkeys made three-dimensional hand movements to reach for small pieces of food randomly placed at four different positions on a tray. Cortical recordings were obtained while the two subjects were trained and tested on both tasks (Fig. 1a).

Figure 1b and c illustrate samples of the raw neuronal data obtained while the animals performed task 1. In both monkeys, coherence analysis^{19–21} revealed that the activity of most single neurons from each of the simultaneously recorded cortical areas was significantly correlated with both one-dimensional (Fig. 2a and b) and three-dimensional hand trajectories, although the degree and frequency range of these correlations varied considerably within and between cortical areas. We then investigated whether both linear^{19–21} and artificial neural network (ANN)^{22,23} algorithms could be used to predict hand position in real time. For one-dimensional movements, we observed that both algorithms yielded highly significant real-time predictions in both monkeys (Fig. 2c and d). These results were obtained in spite of the fact that the trajectories were quite complex, involving different starting positions, as well as movements at different velocities. For example, in the session represented in Fig. 2c, the activity of 27 PMd, 26 MI, 28 PP, and 19 ipsilateral MI/PMd neurons in monkey 1 allowed us to achieve an average correlation coefficient of 0.61 between the observed and predicted hand position (60-minute session, range 0.50–0.71, linear model; 0.45–0.73, ANN; $P < 0.001^{19–21}$). Figure 2d illustrates similar real-time results obtained by using a smaller sample of neurons (8 PMd and 27 MI) in monkey 2 (average $r = 0.72$, range 0.47–0.79, linear model; average $r = 0.66$, range 0.42–0.71, ANN, $P < 0.001$). No large differences in fitting accuracy were observed between linear and ANN algorithms in either animal

(Fig. 2c and d, linear prediction shown as green line, ANN as red line). As shown in Fig. 2e (monkey 1) and Fig. 2f (monkey 2), the performance of both algorithms improved in the first few minutes of recordings and then reached an asymptotic level that was maintained throughout the experiment. In both monkeys, highly significant predictions of hand movement trajectories were obtained for several months.

To reduce the influence of dynamic changes in the coupling between neuronal activity and movements and other non-stationary influences in our real-time predictions, both linear and ANN models were continuously updated throughout the recording sessions. This approach significantly improved the prediction of hand trajectories. For example, when predicting the last 10 minutes of 50–100-minute sessions, the adaptive algorithm performed 55% (20 sessions, median) better than a fixed model based on the initial 10 minutes, or 20% better than a model based on the 30–40-minute segment of the session.

Because accurate hand trajectory predictions were achieved early on in each recording session and remained stable for long periods of time, we were able to use brain-derived signals to control the movements of robotic devices (Phantom, SensAble Technologies)²⁴ in real time (Fig. 2g). In addition, we were able to broadcast these motor control signals to multiple computer clients by using a regular Internet communication protocol (TCP/IP, Fig. 1a) and control two distinct robots simultaneously: one at Duke University (Fig. 2g, blue line) and one at MIT (Fig. 2g, red line).

Next, we investigated whether the same cortical ensemble activity and models could be used to predict the complex sequences of three-dimensional hand movements used by primates in a food-reaching task (task 2). These movements involved four phases:

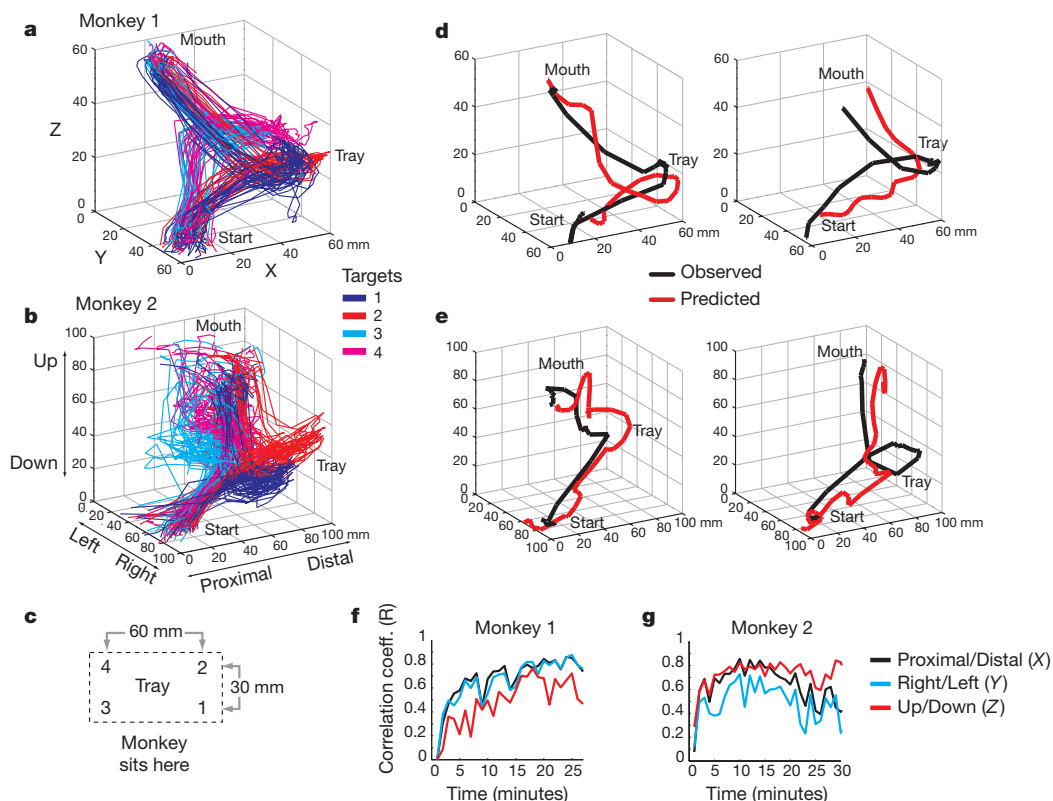


Figure 3 Real-time prediction of 3D hand movements. **a, b**, 3D hand movement trajectories produced by monkey 1 (**a**) and 2 (**b**) during single experimental sessions. **c**, schematic diagram of the four possible target locations in the food reaching task. **d, e**, samples of observed (black line) and real-time predicted (red line) 3D hand

movement for monkey 1 (**d**) and 2 (**e**). **f, g**, Correlation coefficient variation for x (black line), y (blue line) and z (red line) dimensions of predicted 3D hand movements using the linear model.

reaching for the food, grasping the food, bringing the food to the mouth and returning to the start position (Fig. 3a and b). Because these animals were not overtrained, their movement trajectories were highly variable. For example, in the session represented in Fig. 3a (monkey 1) the dispersion of hand trajectories was 7.0 by 7.5 by 6.0 cm (or 315 cm³); in Fig. 3b (monkey 2) the dispersion was even bigger, 11.5 by 10.5 by 10.0 cm (or 1,207.5 cm³). Nonetheless, in both animals, the same linear and ANN models described above provided accurate predictions of three-dimensional hand trajectories in four different directions during 25–60-minute experimental sessions (60–208 trials). Figures 3d and e illustrate several examples of observed (black lines) and predicted (red lines) sequences of three-dimensional movements produced by monkey 1 (Fig. 3d) and 2 (Fig. 3e). After initial improvements, the predictions of three-dimensional hand trajectories reached asymptotic levels that were maintained throughout the experiments (Fig. 3f and g). The three-dimensional predictions were comparable to those obtained for one-dimensional movements (monkey 1: $r = 0.74$, 0.72 and 0.56 for the x -, y - and z -dimensions, respectively; monkey 2: $r = 0.70$, 0.54 and 0.77; 20-minute averages).

Further demonstration of the robustness of our real-time approach was obtained by investigating how well model parameters obtained for one set of hand movements could be used to predict hand trajectories to other directions. For example, by training our linear model only with hand movements directed to targets on the right (targets 1 and 2) we were able to predict accurately hand trajectories to targets on the left (targets 3 and 4). The same was true for the reverse case, that is, using parameters derived from left

movements to predict movements to the right (monkey 1, $r = 0.80$, 0.70 and 0.67 for the x -, y - and z - dimensions; monkey 2, $r = 0.68$, 0.53 and 0.81; averages for both conditions). Predictions of distal (targets 2 and 4) movements by training the model only with proximal (targets 3 and 1) hand trajectories and vice versa were comparably accurate (monkey 1, $r = 0.81$, 0.71 and 0.74 for the x -, y - and z -dimensions; monkey 2, $r = 0.69$, 0.63 and 0.79; averages).

We next analysed how each of the 2–4 different cortical areas contributed to the prediction of one-dimensional movements by calculating the average effect of removing individual neurons, one at a time, from the neuronal population used in each real-time session. This neuron-dropping analysis was carried out independently for each of the cortical areas, as well as for the combination of all of them. We found that hyperbolic functions could fit (r range 0.996–0.9996) all curves that resulted from the neuron-dropping analysis, using both the linear and ANN models in both animals. Figure 4a–e illustrates typical neuron-dropping curves and the corresponding hyperbolic fits. Extrapolations of the hyperbolic curves (Fig. 4f) revealed that 90% correct real-time prediction of one-dimensional movements could be theoretically achieved by applying the linear model to either 480 ± 65.7 PMd neurons, 666 ± 83.0 MI, 629 ± 64.2 PP or $1,195 \pm 142$ ipsilateral MI/PMd neurons in monkey 1 (average and s.e.m., 10 sessions). In monkey 2, the same level of accuracy would require either 376 ± 42.3 PMd neurons or 869 ± 127.4 MI neurons (Fig. 4g). Thus, in both monkeys significantly fewer PMd (red) neurons would theoretically be required to achieve the same level (90%) of one-dimensional hand movement prediction accuracy ($P < 0.001$, Wilcoxon), that is, on average, PMd neurons provided the highest contribution to the predictions. MI (light blue) and PP (dark blue) ensembles provided comparably lower contributions, whereas neurons located in the ipsilateral MI cortex accounted for the lowest amount of variance (yellow line). When all recorded cortical neurons were combined, the extrapolation of the hyperbolic functions produced identical theoretical estimates for 90% prediction accuracy in both monkeys (monkey 1, 625 ± 64 neurons; monkey 2, 619 ± 73 neurons, average and s.e.m.; $P = 0.70$, Wilcoxon, not significant).

These results are consistent with the hypothesis that motor control signals for arm movements appear concurrently in large areas of the frontal and parietal cortices^{5,6}, and that, in theory, each of these cortical areas individually could be used to generate hand trajectory signals in real time. However, the differences in estimated neuronal sample required to predict hand trajectories using a single cortical area probably reflect the functional specializations of these regions. Thus, the differences observed here are consistent with previous observations that PP and MI activity are influenced by motor parameters other than hand position (for example, visual information in the case of PP^{8,25}, or information related to the motor periphery in the case of MI^{26,27}). The relative contributions of these cortical areas may possibly also change according to such factors as the demands of the particular motor task, training level or previous motor experience^{9,15}.

In conclusion, we demonstrated that simultaneously recorded neural ensemble activity, derived from multiple cortical areas, can be used for the generation of both one-dimensional and three-dimensional signals to control robot movements in real time. Contrary to previous off-line algorithms^{5,11,12}, our real-time approach did not make any a priori assumptions about either the physiological properties (for example, shape of the tuning curve) of the single neurons, or the homogeneity of the neuronal population sample. Instead, by using random samples of cortical neurons, we obtained accurate real-time predictions of both one-dimensional and three-dimensional arm movements. In this context, our findings support the notion that motor signals derived from ensembles of cortical neurons could be used for long-term control of prosthetic limb movements^{28,29}. We have shown that chronically implanted microwires can yield reliable recordings in primates for at least 24

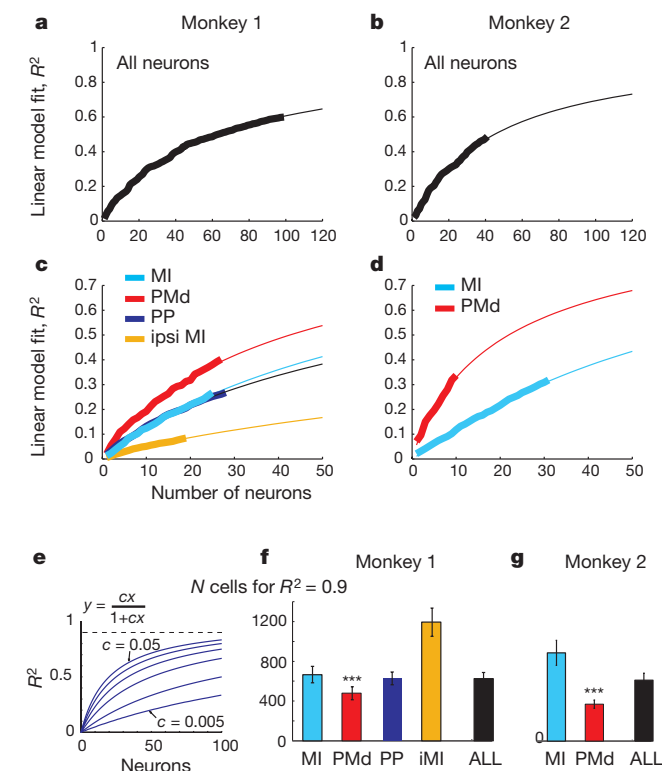


Figure 4 Neuron-dropping analysis. **a, b**, Neuron-dropping curves (thick lines) obtained in a single session for all cortical neurons in monkey 1 (**a**) and 2 (**b**) are precisely fitted by hyperbolic functions (thin lines). **c, d**, Neuron-dropping curves (thick lines) and corresponding hyperbolic functions (thin lines) for each cortical area for monkey 1 (**c**) and 2 (**d**). **e**, Range of fitted hyperbolic functions, x is the number of neurons and c is the fitted constant. **f, g**, Estimated number of neurons per cortical area required to reach an R^2 of 0.9 using the hyperbolic functions in monkey 1 (**f**) and 2 (**g**). Lower values represent higher mean contribution per neuron. Three asterisks, $P < 0.001$.

months; this suggests that a combination of denser multi-wire arrays with implantable integrated circuits, designed to handle all real-time signal processing and mathematical analysis, could one day form the basis of a brain-machine interface for allowing paralysed patients to control voluntarily the movements of prosthetic limbs³⁰. □

Methods

For more detailed methods see Supplementary Information.

Electrophysiological recordings and behavioural tasks

Multiple microelectrode arrays (NBLABS, Dennison) containing 16–32 50- μ m diameter microwires were chronically implanted in different cortical areas^{14–16} in two owl monkeys. Stereotaxic coordinates, published microstimulation maps^{17,18} and intra-operative neural mapping were used to locate the premotor, primary motor and posterior parietal cortical areas. A many-neuron acquisition processor (MNAP, Plexon) was used to acquire and distinguish activity from single neurons¹⁶. Both monkeys were trained and tested on two behavioural tasks. In the first task, the subjects were trained to centre a manipulandum for a variable time and then to move it either to left or right targets in response to a visual cue in order to receive a juice reward. The position of the manipulandum was recorded continuously with a potentiometer. In the second task, the monkeys were trained to place their right hands on a small platform, located waist high and next to their bodies. When an opaque barrier was lifted, the monkeys reached out and grabbed a small piece of fruit from one of four fixed target locations on a tray mounted in front of the platform. In both tasks, the location and orientation of the wrist in three-dimensional space was continuously recorded using a plastic strip containing multiple fibre optic sensors (Shape Tape, Measurand) attached to the right arms of the monkeys. Analogue signals from the strip were sampled at 200 Hz. All sessions were also videotaped.

Data analysis

Predictions of hand position based on simultaneous neural ensemble firing were obtained by applying both a linear model and ANNs to these data. In the linear model, the relationship between the neuronal discharges in the matrix $\mathbf{X}(t)$, and hand position (1D or 3D) in $\mathbf{Y}(t)$ is

$$\mathbf{Y}(t) = \mathbf{b} + \sum_{u=-m}^n \mathbf{a}(u)\mathbf{X}(t-u) + \epsilon(t)$$

where $\mathbf{b}$ are the Y-intercepts in the regression, and $\mathbf{a}$ is a set of impulse response functions describing the weights required for the fitting as function of time lag u . $\epsilon(t)$ are the residual errors. Hence, in this model the series in $\mathbf{X}$ (that is, neuronal firing in time) are convolved with the weight functions $\mathbf{a}$ so that the sum of these convolutions plus $\mathbf{b}$ approximates $\mathbf{Y}$ (hand trajectory)^{19–21}. Offline analysis was first performed to test the validity of this model. Y-intercepts and impulse response functions with a resolution of 5 ms were calculated using a frequency-domain approach^{19,21}. Real-time prediction was achieved by using a slightly modified linear model. Neural data representing potential feedback information from movements were not included in the model. Each neuron's discharges were counted in 100-ms bins. Y-intercepts and weights were calculated using standard linear regression techniques¹⁹. The performance of this simplified real-time model was only very slightly inferior to the complete model outlined above.

Our strategy for applying ANNs is described elsewhere^{14,23}. Here, we employed the same data structure described for the linear model. Several feed-forward ANNs were evaluated offline. The best results were obtained using one hidden layer with 15–20 units, linear output units, the Powell-Beale conjugate gradient training algorithm²², and an early stopping rule. Predictions obtained with ANNs were comparable to those for the linear algorithm.

During real-time prediction of hand position, up to 10 minutes of data were first collected to fit the linear and ANN models. The resulting models were then sent back to the data acquisition computer for instantaneous and continuous prediction of hand position using the neuronal ensemble data acquired in real time. During the remainder of the experimental session, both models continued to be calculated repeatedly, using the most recently recorded 10 minutes of data. All calculations were performed by software designed in MATLAB (The Mathworks, Natick, Massachusetts, USA).

Real-time control of robot arms

Predictions of hand positions were broadcasted by a TCP/IP server to one or more computer clients. One of these clients controlled the movements of a robot arm (Phantom model A1.0; Phantom, SensAble Technologies)²⁴ locally at Duke University, and another client was used to control simultaneously the movements of a remote robot arm (Phantom Desktop) at MIT using the Internet. Signals describing the position of the robot arm in space were recorded on each client machine.

Neuron-dropping analysis

The relation between ensemble size and the goodness of fit for our models was analysed offline by randomly removing single neurons, one at a time, and fitting the models again using the remaining population until only one neuron was left. This procedure was repeated 25–30 times for each ensemble, so that an average curve was obtained to describe

R^2 as a function of ensemble size. Neuron-dropping curves were obtained for all cortical areas (separately or combined) in each monkey, and simple hyperbolic functions were used to fit them with very high accuracy. Next, we used these functions to estimate the ensemble size that would theoretically be required to attain a R^2 of 0.9.

Received 10 July; accepted 6 October 2000.

- Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. L. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci.* **2**, 664–670 (1999).
- Evarts, E. V. Pyramidal tract activity associated with a conditioned hand movement in the monkey. *J. Neurophysiol.* **29**, 1011–1027 (1966).
- Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. Posterior parietal association cortex of the monkey: command functions for operation within extrapersonal space. *J. Neurophysiol.* **38**, 871–908 (1975).
- Fetz, E. E. & Cheney, P. D. Muscle fields of primate corticomotoneuronal cells. *J. Physiol. (Paris)* **74**, 239–245 (1978).
- Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
- Weinrich, M. & Wise, S. P. The premotor cortex of the monkey. *J. Neurosci.* **2**, 1329–1345 (1982).
- Wise, S. P., Boussaoud, D., Johnson, P. B. & Caminiti, R. Premotor and parietal cortex: corticocortical connectivity and combinatorial computations. *Annu. Rev. Neurosci.* **20**, 25–42 (1997).
- Batista, A. P., Buneo, C. A., Snyder, L. H. & Andersen, R. A. Reach plans in eye-centered coordinates. *Science* **285**, 257–260 (1999).
- Mitz, A. R., Godschalk, M. & Wise, S. P. Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations. *J. Neurosci.* **11**, 1855–1872 (1991).
- Humphrey, D. R., Schmidt, E. M. & Thompson, W. D. Predicting measures of motor performance from multiple cortical spike trains. *Science* **170**, 758–762 (1970).
- Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. Neuronal population coding of movement direction. *Science* **233**, 1416–1419 (1986).
- Schwartz, A. Direct cortical representation of drawing. *Science* **265**, 540–542 (1994).
- Fetz, E. E. & Cheney, P. D. Postspike facilitation of forelimb muscle activity by primate corticomotoneuronal cells. *J. Neurophysiol.* **44**, 751–772 (1980).
- Nicolelis, M. A. L. et al. Simultaneous representation of tactile information by distinct primate cortical areas rely on different encoding strategies. *Nature Neurosci.* **1**, 621–630 (1998).
- Laubach, M., Wessberg, J. & Nicolelis, M. A. L. Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. *Nature* **405**, 567–571 (2000).
- Nicolelis, M. A. L., Ghazanfar, A. A., Faggini, B. M., Votaw, S. & Oliveira, L. M. Reconstructing the engram: simultaneous, multisite, many single neuron recordings. *Neuron* **18**, 529–537 (1997).
- Stepniowska, I., Preuss, T. M. & Kaas, J. H. Architectonics, somatotopic organization, and ipsilateral cortical connections of the primary motor area (M1) of owl monkeys. *J. Comp. Neurol.* **330**, 238–271 (1993).
- Preuss, T. M., Stepniowska, I. & Kaas, J. H. Movement representation in the dorsal and ventral premotor areas of owl monkeys: a microstimulation study. *J. Comp. Neurol.* **371**, 649–676 (1996).
- Brillinger, D. R. *Time Series. Data Analysis and Theory* (Holden-Day, San Francisco, 1981).
- Bendat, J. S. & Piersol, A. G. *Random Data. Analysis and Measurement Procedures* (Wiley, New York, 1986).
- Halliday, D. M. et al. The Fourier approach to the analysis of mixed time series/point process data. Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Prog. Biophys. Mol. Biol.* **64**, 237–278 (1995).
- Powell, M. J. D. Restart procedures for the conjugate gradient method. *Math. Program.* **12**, 241–254 (1977).
- Ghazanfar, A. A., Stambaugh, C. R. & Nicolelis, M. A. L. Encoding of tactile stimulus location by somatosensory thalamocortical ensembles. *J. Neurosci.* **20**, 3761–3775 (2000).
- Salisbury, J. K. & Srinivasan, M. A. Phantom-based haptic interaction with virtual objects. *IEEE Comput. Graph. Appl.* **17**, 6–10 (1997).
- Ferraina, S. et al. Combination of hand and gaze signals during reaching: Activity in parietal area 7m of the monkey. *J. Neurophysiol.* **77**, 1034–1038 (1997).
- Mussa-Ivaldi, F. A. Do neurons in the motor cortex encode movement direction? An alternative hypothesis. *Neurosci. Lett.* **91**, 106–111 (1988).
- Scott, S. H., Sergio, L. E. & Kalaska, J. F. Reaching movements with similar hand paths but different arm orientations. II. Activity of individual cells in dorsal premotor cortex and parietal area 5. *J. Neurophysiol.* **78**, 2413–2426 (1997).
- Schmidt, E. M. Single neuron recording from motor cortex as a possible source of signals for control of external devices. *Ann. Biomed. Eng.* **8**, 339–349 (1980).
- Kennedy, P. R. & Bakay, R. A. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport* **9**, 1707–1711 (1998).
- Nicolelis, M. A. L. Hybrid brain-machine interfaces for translating thoughts into action. *Nature* (in the press).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by grants from the National Institutes of Health and DARPA-ONR to M.A.L.N., J.K.C. and M.A.S., and the National Science Foundation to M.A.L.N. J.W. was supported by The Swedish Foundation for International Cooperation in Research and Higher Education, and the Swedish Medical Research Council. J.K., P.B. and M.L. were supported by NIH postdoctoral fellowships.

Correspondence and requests for materials should be addressed to M.A.L.N. (e-mail: nicoleli@neuro.duke.edu).

A regulator of transcriptional elongation controls vertebrate neuronal development

Su Guo[†], Yuki Yamaguchi[‡], Sarah Schilbach^{*}, Tadashi Wada[‡], James Lee^{*}, Audrey Goddard^{*}, Dorothy French[§], Hiroshi Handa[‡] & Arnon Rosenthal^{*}

^{*} Departments of Molecular Biology and Pathology[§], Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA

[‡] Frontier Collaborative Research Center (FCRC) and Department of Biological Information Tokyo Institute of Technology 4259 Nagatsuta, Yokohama 226-8501, Japan

The development of distinct vertebrate neurons is defined by the unique profiles of genes that neurons express. It is accepted that neural genes are regulated at the point of transcription initiation, but the role of messenger RNA elongation in neural gene regulation has not been examined^{1–3}. Here we describe the mutant *foggy*, identified in a genetic screen for mutations that affect neuronal development in zebrafish⁴, that displayed a reduction of dopamine-containing neurons and a corresponding surplus of serotonin-containing neurons in the hypothalamus. Positional cloning disclosed that Foggy is a brain-enriched nuclear protein that is structurally related to the transcription elongation factor Spt5 (refs 5–12). Foggy is not part of the basic transcription apparatus but a phosphorylation-dependent, dual regulator of transcription elongation. The mutation disrupts its repressive but not its stimulatory activity. Our results provide molecular, genetic and biochemical evidence that negative regulators of transcription elongation control key aspects of neuronal development.

Zebrafish with the *foggy* mutation have a typical anatomy (Fig. 1a, b) and do not suffer excessive cell death or deficit in cell proliferation (as judged by TUNEL (terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick end labelling) staining and 5-bromo-2'-deoxyuridine (BrdU) incorporation; see Supplementary Information). Despite that, they fail to develop a normal complement of tyrosine hydroxylase (TH) positive dopamine (DA) neurons in the hypothalamus (~17 ± 3 in wild type, ~8 ± 2 in *foggy*; mean ± s.d.), and the few remaining neurons expressed reduced levels of TH (Fig. 1c, d). Unexpectedly, the number of hypothalamic serotonin (5HT) neurons that developed in close proximity to the DA neurons more than doubled from ~6 ± 1 to ~15 ± 2 (Fig. 1c, d). Another group of DA neurons that irreversibly failed to develop was the interneurons in the retina (Fig. 1e, f). Five other classes of retinal neurons that share progenitors with the DA interneurons¹³, including the GABA⁺ amacrine interneurons and the rod and cone photoreceptors that normally express the red, blue and ultraviolet opsin, rhodopsin and Fret43 (Fig. 1g–j; Supplementary Information), were also severely reduced in number. In addition, the axonal projections of these neurons, which constitute the inner and outer plexiform layers of the retina, were mainly absent (Fig. 1k, l). In contrast, the number of ZN-8⁺ ganglion neurons and their axons that constitute the optic nerve appeared normal, or had perhaps increased (Fig. 1m, n). Despite the deficit of mature photoreceptor and amacrine neurons, undifferentiated cells were present in normal numbers in the inner nuclear and photoreceptor layers (Fig. 1k, l; see cell counting in figure legend and BrdU incorporation in Supplementary Information).

Other cell types, including TH⁺ arch-associated cells, 3A10⁺ Mauthner and 5HT⁺ 5HT neurons in the hindbrain, GABA⁺

interneurons and islet1⁺ motoneurons in the spinal cord, appear normal. Also, the expression patterns of hypothalamic and retinal regional markers, as well as other regulatory and structural proteins including the secreted proteins DeltaA, DeltaB, DeltaC, DeltaD, Notch, Notch-1B, Shh and FGF8 were unchanged. Also, the transcription factors Grg1, Grg2, Krox20, Otx2, Zash-1a, Islet1, Nurr1, NeuroD, Zrx3, Nkx2.2 and Neurogenin, the receptors Ret and Rack, and the cellular markers GFAP, Hu, 3A10, Zn8 and γ -crystallin were

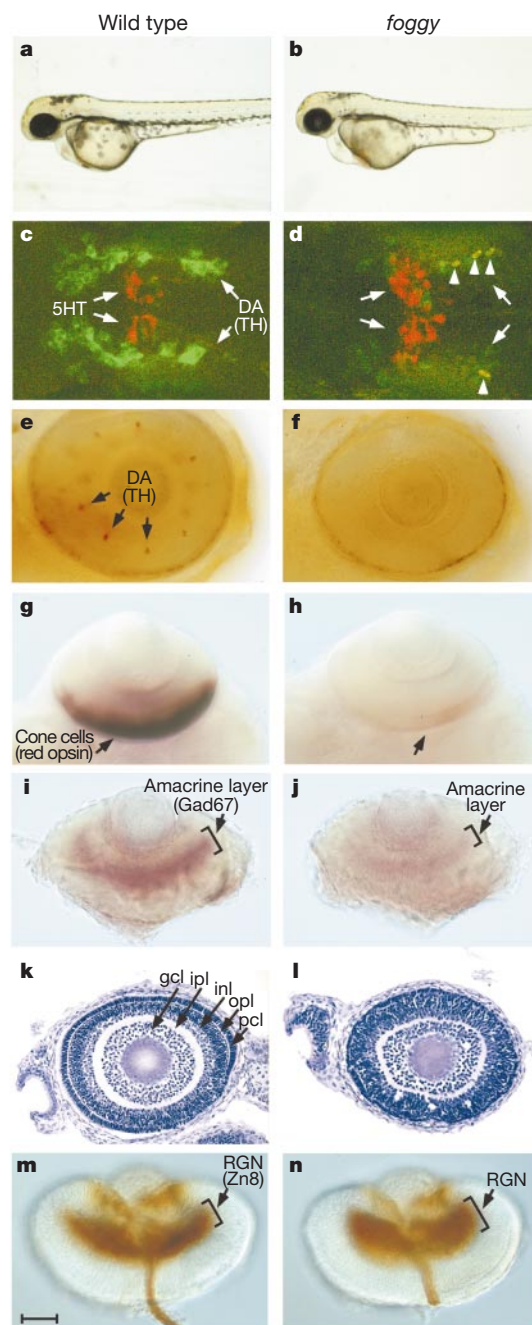


Figure 1 The *foggy* mutation disrupts neuronal development. **a, b**, Morphology of 2-day embryos. **c, d**, Reduction in DA neurons and increase in 5HT neurons in *foggy*. **e–j**, The numbers of DA amacrine neurons (**e, f**), red opsin⁺ photoreceptor neurons (**g, h**) and GAD67⁺ amacrine neurons (**i, j**) are reduced in the mutant retina. **k, l**, Histological analysis of eyes at 72 h. The inner nuclear layer contains 7.4 (wild-type) and 7.6 (mutant) cells respectively along the retina diameter. **m, n**, Normal complement of Zn-8⁺ ganglion neurons. gcl, ganglion cell layer; inl, inner nuclear layer; ipl, inner plexiform layer; opl, outer plexiform layer; pcl, photoreceptor layer; RGN, retinal ganglion neurons. Scale bar, 300 μ m (**a, b**), 55 μ m (**c–n**).

[†] Present address: Department of Biopharmaceutical Sciences and Program in Human Genetics, University of California, San Francisco, California 94143-0446, USA.

unchanged (see Supplementary Information). The *foggy* mutant dies at around day 4 from deficit in blood circulation that appears to be a secondary consequence of abnormal neuronal function (data not shown). Together, these findings indicate that *foggy* is mutated in a gene that regulates the development of distinct neuronal classes.

To identify the *foggy* gene, we used an automated amplified fragment length polymorphism (AFLP) DNA fingerprinting technology¹⁴ and chromosomal walking (see Supplementary Information). Bacterial artificial chromosomes (BACs) that covered the *foggy* locus were injected individually into *foggy* mutant embryos at the 1–2 cell-stage, and their ability to rescue the *foggy* phenotype was examined. BAC B108J11 and BAC B35I11 were able to restore normal neuronal development to homozygous *foggy* mutants (Fig. 2). Shotgun sequencing of the entire BAC B108J11 and a screen of complementary DNA library using BAC B108J11 as a probe allowed us to identify two transcription units on this BAC. Injection of full-length cDNA corresponding to the proximally located transcription unit under the control of zebrafish β -actin promoter¹⁵ further showed that it can rescue the *foggy* mutant embryos.

The rescuing cDNA encodes a protein of 1,084 amino acids. The protein comprises an amino-terminal acidic region and a central area with four NusG/KOW homology domains that are found in a bacterial elongation factor^{16–18}. The carboxy terminus displays three types of hexapeptide repeat: four XTPXYG, two QTPLHD and two NPQTPG (Fig. 3a). A database search suggested that Foggy belongs to an evolutionarily conserved family, whose founding member is the yeast nuclear protein Spt5 (ref. 5; Fig. 3b). Mutations in the yeast Spt5 are deleterious and associated with changes in chromatin structure and transcription elongation of a subset of genes^{5–7}. Although the physiological role of Spt5 has never been elucidated in multicellular organisms, *in vitro* biochemical studies have shown that it binds to¹⁰, and inhibits transcription elongation by RNA polymerase II (Pol II) in a phosphorylation-dependent manner¹¹. When Pol II is phosphorylated at its C terminus by the serine/threonine kinase p-TEFb, it is resistant to the inhibitory activity of Spt5. When the p-TEFb kinase is in an inactive state, however, Pol II becomes hypophosphorylated and susceptible to negative regulation by Spt5. To elicit its inhibitory activity, Spt5 must interact with the nuclear protein Spt4 of relative molecular mass 14,000 (M_r 14K) (ref. 11), with a negative elongation factor protein complex⁹ and possibly with the chromatin-binding protein Spt6 (ref. 6) that is

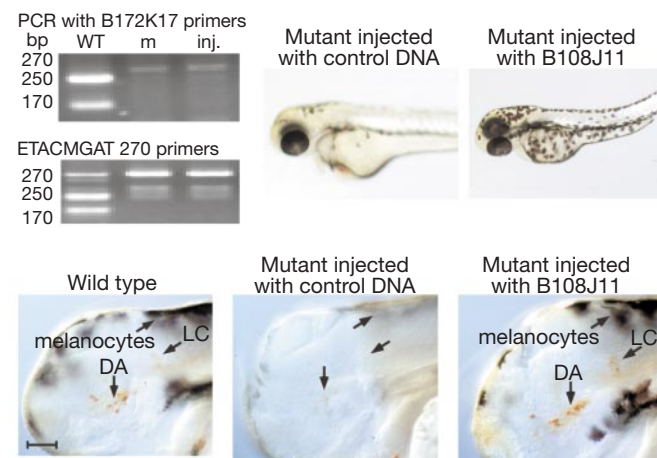


Figure 2 Identification of the *foggy* gene. BAC B108J11 rescues the *foggy* mutant phenotype. Embryos from *foggy*^{m806} heterozygous animals were injected with purified BAC DNA, allowed to develop to 2-day old, and immunostained with TH antibody (middle and right). Wild type is shown bottom left. Scale bar, 100 μ m. Rescued embryos were genotyped by closely linked polymorphic markers and subsequent sequencing (top left).

a 1 MSDESDSDSDNQSERSSSEAEVEEENEEEEEQGSVAGSDKAE
44 EGEDLEDEEEYDEEEEDDDRPKKARHGGLFLEADVDDEYE
87 DEDPWEDGAEDILEKEEAESVNLHDVLDHSDGSRRLQNLWR
130 DSREELAGEYYMRKYAKSSGGEHFYGGSEDLSDITQQQLPG
173 VKDPNLWTVCKIGEEERATAISLMRFVAYQCTDTPLQIKSVV
216 APEHVKGYYVEAYKQTHVKAIEGVNLRMGFWNQMPVPIKE
259 MTDVLKVVKEVTNLKPKSWRLKRGLYKDDIAQVDYVEPSQNT
302 ISLKMIPRIDLDRIKARMSMKDWFAKKKKFKRPPQRLFDAEKI
345 RSLGGEVSHDGDGMIFANRYSRKGLFKSFAMSAVITEGVKP
388 TLSELEKFEDQPEGIDLEVVTETTGKEREHNLQAGDNVEVCEG
431 ELINLOGKILSVGDNKITIMPKHEDLKDPLEPPAHELRLKYFRM
474 GDHVKVIAGRYEGDTGLIVRVEENFVILFSDLTMHELKVLPRD
517 LQLCSETASGVDAGGQHEWGLVQLDPQTVGVIVRLERETFOV
560 LNMHGKVLTVRHQAVNRKDNRFVALDSEQNNIHVKDVKVI
603 DGPHSGREGEIRHIFRGFAFLHCKKLVENGGMFVCKARHLVA
646 GGSKPRDVTNFTVGGFAPMSPRISSPMHPGGGGQPGGGGGG
689 GGMGRGRGRDNDLIGQTVRISOGPKYIGVVKDATESTAR
732 VELHSTCQTISVDRQLRTTVGGKERQGRSSTHLRTPMYGSQTP
775 IYGTGSRTPMYGSQTPPLHDGSRTPHYGSQTPPLHDGSRTPGQSG
818 AWDPNPNTPSRPDDEYEFAYDDEPSPSPQGYGGTNPQTPGY
861 PEVPSPQVNPQYNPQTPGTPAMYNTDQYSPYAAPSQGSYQPS
904 PSPQSYHQVAPSPVGYQNTSPASYPHTPSPMAYQASPSPSV
947 GYSPMTPGAPSPGGYNPHTPGSNIDQASNDVWTTDIMVRVKDT
990 FLGGVINQGTGIIRSVTGGMCSVFLQDTEKVVSISSSEHLEPVT
1033 PTKNNKVKVILGEDREATGVLLSIDGEDGIVRMELDEQLKILN
1076 LRFLGKLEV

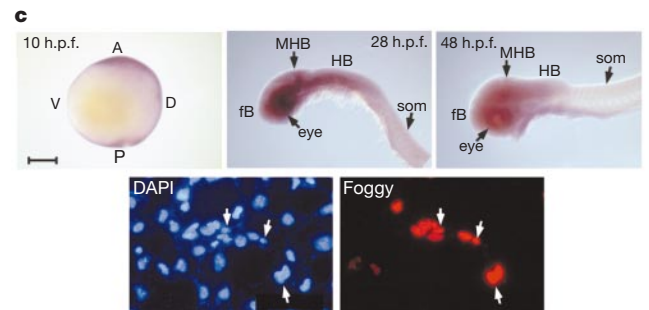
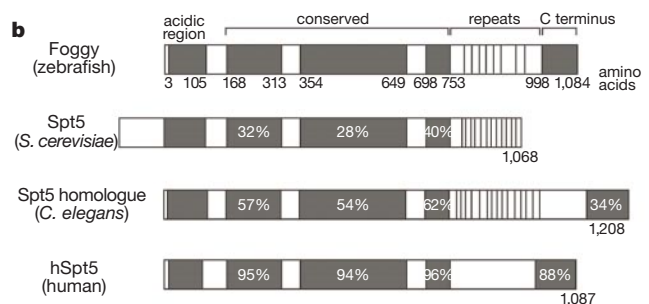


Figure 3 Sequence and domain structure of the Foggy protein. **a**, Primary sequence. The N-terminal acidic region is underlined. Four KOW motifs are underlined and labelled 1–4. Two putative nuclear localization signals are boxed. Hexapeptide repeats are underlined twice. Conserved residues in the C-terminal domain are marked with asterisks. The single amino-acid mutation detected in *foggy*^{m806} is marked. **b**, Comparison of Foggy with yeast, *Caenorhabditis elegans* and human Spt5. **c**, Whole-mount *in situ* hybridization of 10 h.p.f., 28 h.p.f. and 2-day embryos and Cos-7 cells transfected with Flag-tagged wild-type *foggy* construct, visualized by DAPI (blue) and Flag antibody staining (red). Scale bar, 160 μ m (top); 40 μ m (bottom). FB, forebrain; HB, hindbrain; MHB, mid-hindbrain boundary; som, somites.

associated with the Notch signalling system¹⁹. Even though Spt5 inhibits elongation by hypophosphorylated Pol II at normal nucleotide concentrations, at low nucleotide concentrations it acts as an elongation stimulator^{7,11}, suggesting that it is a dual transcriptional regulator whose activity is under the control of signalling molecules and the environment.

The *foggy* mutant carries a single nucleotide substitution from T to A, which changes the amino-acid valine 1,012 to aspartic acid (V1012D; see Supplementary Information). Valine 1,012 is an evolutionarily invariant amino acid that is located in the conserved

C-terminal domain (Fig. 3b). Unlike its wild-type counterpart, a construct carrying the mutation failed to rescue the *foggy* mutant phenotype (data not shown). *foggy* messenger RNA is detected in the neural plate of the tailbud-stage embryo (~10 hours post fertilization (h.p.f.)). At ~28 h.p.f., it is found predominantly in the developing brain and at lower levels elsewhere. By 48 h.p.f., the number of *foggy* transcripts is decreased throughout the organism (Fig. 3c). A similar expression pattern is detected in the *foggy* mutant embryos (data not shown), suggesting that the mutation does not affect *foggy* mRNA stability. Both wild-type and the V1012D mutant forms of Foggy are detected in the cell nucleus, indicating that Foggy is a nuclear protein and that the mutation does not disrupt its subcellular localization (Fig. 3c; and data not shown). The specificity of the *foggy* phenotype combined with the wide distribution of the *foggy* mRNA suggests that Foggy activity is regulated by post-transcriptional modifications.

We next tested whether *foggy* is a regulator of transcription elongation, and if so whether the V1012D mutation disrupts its transcriptional activity. When we added recombinant wild-type or mutant Foggy proteins to a transcription elongation complex without the drug 6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB), which inactivates the p-TEFb kinase¹¹, there was no change in the rate of transcription elongation. Likewise, transcription elongation progressed normally in the presence of DRB alone (Fig. 4). Thus, Foggy is not part of the basal Pol II transcription elongation machinery, and in the absence of Foggy phosphorylation of Pol II by p-TEFb is not required for transcription elongation. In contrast, when we added wild-type Foggy together with DRB, there was a roughly fourfold inhibition in transcription elongation by the hypophosphorylated Pol II (Fig. 4b, lanes 5, 6). Similar inhibition was induced by the mammalian recombinant Spt5 (Fig. 4b, lanes 1, 2). The recombinant mutant Foggy protein was completely inactive and failed to inhibit elongation in the presence of DRB, even when added at a ninefold higher concentration (Fig. 4b, lanes 7–12).

To measure any stimulatory activity of wild-type and mutant Foggy, we examined their effect on transcription elongation at low nucleotide concentrations. When we added recombinant wild-type Foggy or human recombinant Spt5 to a transcription elongation with low nucleotide concentrations, each protein led to a similar increase in the efficiency of transcription elongation (Fig. 4c, d, lanes 9–12; and data not shown). Unexpectedly, the mutant Foggy was as efficient as wild-type protein in stimulating elongation at low nucleotide concentrations, increasing the efficiency of elongation by 4.3-fold at 15 min and by 3.6-fold at 20 min (Fig. 4c, d, lanes 13–16). Thus, the V1012D mutant has lost its ability to act as a negative regulator but has retained its ability to stimulate transcription elongation. Together these findings support the hypothesis that Foggy is a phosphorylation-dependent dual regulator of transcription elongation, homologous to Spt5, and that negative rather than positive regulation of transcription elongation by Foggy is essential for the development of distinct neuronal subtypes.

Several regulatory genes whose expression must be downregulated before cell differentiation can take place have been identified. These include the neuron-restrictive silencer factor NRSF/REST^{20,21}, the transcriptional repressors Hes and Id²², the glial POU domain transcription factor Tst-1/Oct6/SCIP²³, the lymphoid zinc-finger factor PRDI/BF1/Blimp²⁴ and the haematopoietic zinc finger transcription factor GATA-2 (ref. 25). Similar genes could be under negative regulation by Foggy/Spt5. Although the full physiological role of regulators of transcription elongation is not known, biochemical studies have revealed the presence of over a dozen such factors^{1–3}, indicating that gene regulation at the transcription elongation level, which is prevalent in prokaryotes, may be a widespread phenomena also in vertebrates. Combinatorial regulation of transcription initiation and elongation would enable versatility in the control of gene expression, and in the generation of diverse cell types with a relatively small number of factors. □

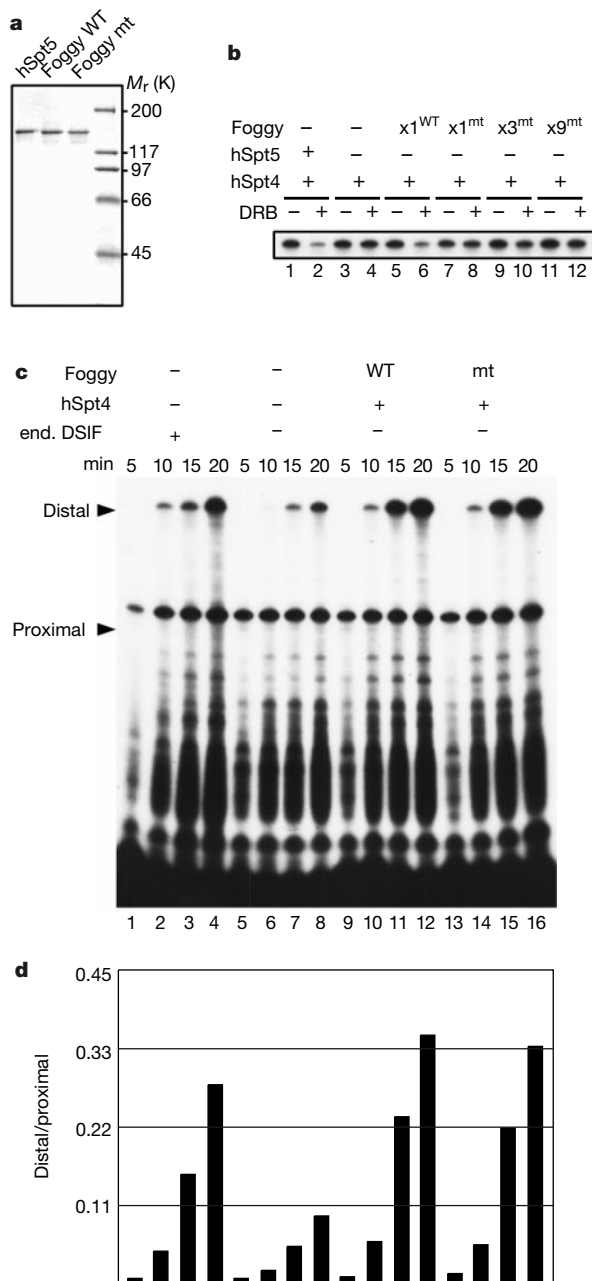


Figure 4 The V1012D mutation abolishes the negative but not positive function of Foggy in transcription elongation. **a**, Recombinant proteins visualized by SDS-PAGE and Coomassie staining. **b**, Depletion/add-back assays using crude HeLa nuclear extract, showing that the mutant Foggy lost its repressive activity. **c**, Depletion/add-back assays under limiting nucleotide concentrations, showing normal stimulatory activity of the mutant Foggy protein. The amounts of two RNA bands were quantified using an FLA2000 image analyser (Fuji), and the promoter distal to proximal ratios were calculated. **d**, Quantitative presentation of the ratio between the distal and proximal G-rich cassettes in **c**.

Methods

Genetic mapping and positional cloning

Heterozygous female fish (AB/EK) carrying the *foggy* mutation were crossed with wild-type male fish from a Tu genetic background. F₂ progeny were identified and separated into wild-type, heterozygous and mutant pools of 40 individuals. DNA from these fish was used for AFLP analysis (see Supplementary Information). We used polymerase chain reaction (PCR) primers corresponding to the two most closely linked AFLP markers to screen BAC (Genome Systems) and yeast artificial chromosome (Research Genetics) libraries and to initiate chromosomal walk. Individual BACs were then injected at a concentration of 40–60 ng μl^{-1} into 1–2-cell stage zebrafish embryos derived from *foggy* heterozygous mating. BAC B108J11, which rescued the *foggy* mutant phenotype, was digested with *Eco*R1, random-primed with p32, and used to screen a 33-h embryonic zebrafish cDNA library. Full-length cDNAs were cloned into the vector containing β -actin promoter, injected at a concentration of 15–30 ng μl^{-1} into embryos, and assayed for rescue. To identify the *foggy* mutation, we used gene-specific primers to amplify genomic DNA from pools of (~5) *foggy* mutant and wild-type sibling embryos. PCR products were directly sequenced using automated cycle sequencers (ABI). cDNAs from mutant and wild-type sibling embryos were synthesized by PCR with reverse transcription (RT-PCR) and sequenced. We used site-directed mutagenesis to introduce the single nucleotide change T to A using reagents from Stratagene.

Transcription assays

A *Nde*I site was created at the 5'-end of the zSpt5 open reading frame by PCR mutagenesis. The *Nde*I (partial)/ *Not*I fragment containing the full-length, wild-type or mutant *foggy* was inserted into *Nde*I–*Bam*HI sites of pET-14b (Novagen), yielding pET-zSpt5 wild-type and pET-zSpt5 mutant. Lysates were prepared from a transformed *Escherichia coli* BL21 (DE3) strain. His₆-zSpt5 proteins were purified from the supernatants by Ni-affinity chromatography, separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and recovered from gel slices. Indicated amounts of recombinant hSpt4 and hSpt5, wild-type or mutant *foggy* were added back to HeLa nuclear extract (2 μl) that was immunodepleted by hSpt4&5 monoclonal antibody. We used pTF3-6C2AT (25 ng) or pSLG402 (125 ng) as a template. To test for stimulatory activity, we used a pSLG402 vector containing a promoter-proximal 85-nucleotide and a promoter-distal 377-nucleotide G-free cassette. The resulting G-free mRNA regions are resistant to RNase T1 digestion, whereas the G-rich linker region is RNase T1 sensitive. After RNase T1 digestion, the two cassettes can be isolated and separated by gel electrophoresis. The molar ratio of the distal cassette, which represents transcription elongation events, to the proximal transcription cassette, which reflects an initiation event, measures the elongation efficiency. The low NTP concentrations were 30 μM ATP, 30 μM GTP, 300 μM CTP and 2.5 μM UTP.

Received 31 May; accepted 17 August 2000.

- Greenblatt, J. RNA polymerase II holoenzyme and transcriptional regulation. *Curr. Opin. Cell Biol.* **9**, 310–319 (1997).
- Shilatifard, A., Conaway, J. W. & Conaway, R. C. Mechanism and regulation of transcriptional elongation and termination by RNA polymerase II. *Curr. Opin. Genet. Dev.* **7**, 199–204 (1997).
- Uptain, S. M., Kane, C. M. & Chamberlain, M. J. Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* **66**, 117–172 (1997).
- Guo, S. *et al.* Mutations in the zebrafish unmask shared regulatory pathways controlling the development of catecholaminergic neurons. *Dev. Biol.* **208**, 473–487 (1999).
- Swanson, M. S., Malone, E. A. & Winston, F. SPT5, an essential gene important for normal transcription in *Saccharomyces cerevisiae* encodes an acidic nuclear protein with a carboxy-terminal repeat. *Mol. Cell Biol.* **11**, 3009–3019 (1991).
- Swanson, M. S. & Winston, F. SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in *Saccharomyces cerevisiae*. *Genetics* **132**, 325–326 (1992).
- Hartzog, G. A., Wada, T., Handa, H. & Winston, F. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev.* **12**, 357–369 (1998).
- Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D. & Hande, H. Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription *in vitro*. *EMBO J.* **17**, 7395–7403 (1998).
- Yamaguchi, Y. *et al.* NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).
- Yamaguchi, Y. *et al.* Structure and function of the human transcription elongation factor DSIF. *J. Biol. Chem.* **274**, 8085–8092 (1999).
- Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
- Yamaguchi, Y., Wada, T. & Handa, H. Interplay between positive and negative elongation factors: drawing a new view of DRB. *Genes Cells* **3**, 9–15 (1998).
- Cepko, C. L. The roles of intrinsic and extrinsic cues and bHLH genes in the determination of retinal cell fates. *Curr. Opin. Neurobiol.* **9**, 37–46 (1999).
- Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414 (1995).
- Higashijima, S., Okamoto, H., Ueno, N., Hotta, Y. & Eguchi, G. High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles of the whole body by using promoters of zebrafish origin. *Dev. Biol.* **192**, 289–299 (1997).
- Sullivan, S. L. & Gottesman, M. E. Requirement for *E. coli* NusG protein in factor-dependent transcription termination. *Cell* **68**, 989–994 (1992).
- Sullivan, S. L., Ward, D. F. & Gottesman, M. E. Effect of *Escherichia coli* nusG function on lambda N-mediated transcription antitermination. *J. Bacteriol.* **174**, 1339–1344 (1992).
- Kypides, N. C., Woese, C. R. & Ouzounis, C. A. KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem.* **21**, 425–426 (1996).
- Hubbard, E. J., Dong, Q., Greenwald, I. Evidence for physical and functional association between EMB-5 and LIN-12 in *Caenorhabditis elegans*. *Science* **273**, 112–115 (1996).

- Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1363 (1995).
- Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nature Genet.* **20**, 136–142 (1998).
- Anderson, D. J. & Jan, Y. N. in *Molecular and Cellular Approaches to Neural Development* (ed. Cowan, W. M.) 26–63 (Oxford Univ. Press, New York, 1997).
- Zorick, T. S., Syroid, D. E., Brown, A., Gridley, T. & Lemke, G. Krox-20 controls SCIP expression, cell cycle exit and susceptibility to apoptosis in developing myelinating Schwann cells. *Development* **126**, 1397–1406 (1999).
- Turner, C. A. Jr, Mack, D. H. & Davis, M. M. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77**, 297–306 (1994).
- Persons, D. A. *et al.* Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood* **93**, 488–499 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank E. Chen and P. Ma for help with the AFLP analysis; Y. Yan and J. Postlethwait for anchoring our AFLP markers to the zebrafish genetic map; A. Greenleaf and D. Price for providing the plasmid pSLG402; M. Hynes, J. Lin, B. Lu M. Tessier Lavigne, B. Barres and S. Wilson for critically reading the manuscript; D. Anderson for helpful information; and J. Ligos, A. Bruce and V. Goodwin for help with graphics. Y.Y. is a JSPS Research Fellow. This work was supported in part by a grant-in-aid for Scientific Research on Priority Areas from the Ministry of Education, Sciences, Sports, and Culture of Japan, and a grant from NEDO to H.H.

Correspondence and requests for materials should be addressed to A.R. (e-mail: ar@gene.com). The GenBank accession number for *foggy* is AF288409.

CIC-5 Cl[−]-channel disruption impairs endocytosis in a mouse model for Dent's disease

Nils Piwon*, Willy Günther*, Michael Schwake, Michael R. Bösl & Thomas J. Jentsch

Zentrum für Molekulare Neurobiologie Hamburg, ZMNH, Universität Hamburg, Martinistrasse 85, D-20246 Hamburg, Germany
* These authors contributed equally to this work.

Dent's disease is an X-linked disorder associated with the urinary loss of low-molecular-weight proteins, phosphate and calcium, which often leads to kidney stones^{1,2}. It is caused by mutations³ in CIC-5, a renal chloride channel^{4,5} that is expressed in endosomes of the proximal tubule^{6,7}. Here we show that disruption of the mouse *clcn5* gene causes proteinuria by strongly reducing apical proximal tubular endocytosis. Both receptor-mediated and fluid-phase endocytosis are affected, and the internalization of the apical transporters NaPi-2 and NHE3 is slowed. At steady state, however, both proteins are redistributed from the plasma membrane to intracellular vesicles. This may be caused by an increased stimulation of luminal parathyroid hormone (PTH) receptors owing to the observed decreased tubular endocytosis of PTH. The rise in luminal PTH concentration should also stimulate the hydroxylation of 25(OH) vitamin D₃ to the active hormone. However, this is counteracted by a urinary loss of the precursor 25(OH) vitamin D₃. The balance between these opposing effects, both of which are secondary to the defect in proximal tubular endocytosis, probably determines whether there will be hypercalciuria and kidney stones.

We disrupted the *clcn5* gene by replacing genomic segments encoding the functionally important stretch between transmembrane domains D2 and D5^{8,9} by a lacZ/neomycin cassette (Fig. 1a). This led to a loss of a normal CIC-5 messenger RNA in *clcn5*[−] mice (Fig. 1b). Western blots using an amino-terminal CIC-5 antibody⁶ indicated that the predicted fusion protein may be unstable (Fig. 1c).

Male *clcn5*^{-/-}, female *clcn5*^{-/-} and female *clcn5*^{+/-} animals grew normally and were fertile. *clcn5*^{-/-} mice did not have kidney stones or nephrocalcinosis even after one year. No changes in kidney morphology and histology were detected. When compared to wild-type (WT) mice, *clcn5*^{-/-} mice produced slightly more urine, which was slightly acidic. *clcn5*^{-/-} mice had no hypercalciuria, but

urinary phosphate excretion was increased by about 50% (see Table in Supplementary Information). *clcn5*^{-/-} mice displayed low-molecular-weight proteinuria. Vitamin D binding protein and retinol binding protein, which is lost into the urine of patients with Dent's disease², were highly elevated in the urine of *clcn5*^{-/-} mice (Fig. S1 in Supplementary Information). Neither proteinuria nor

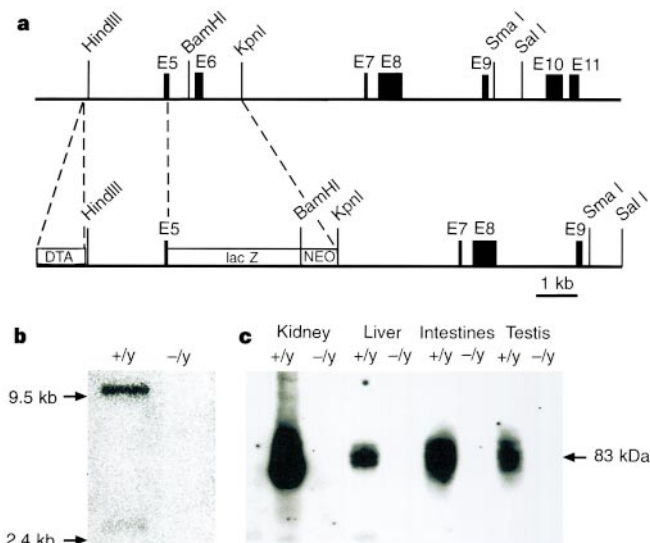


Figure 1 Generation of *clcn5*^{-/-} mice. **a**, Diagram of the targeting vector. Parts of exon 5, exon 6 and the intervening intron are replaced by a lacZ/neomycin cassette. A diphtheria toxin- α gene (DTA) was fused to select against random integration. **b**, Northern analysis of kidney mRNA from *clcn5*^{+/-} and *clcn5*^{-/-} mice. A probe corresponding to exon 6 detects

CIC-5 mRNA only in WT mice. **c**, Western analysis of membranes from various tissues, using the PEP5A1 N-terminal CIC-5 antibody⁶. Together with immunocytochemical results (Figs 2 and 3), it also confirms the lack of cross-reactivity against the related CIC-3 and CIC-4 proteins.

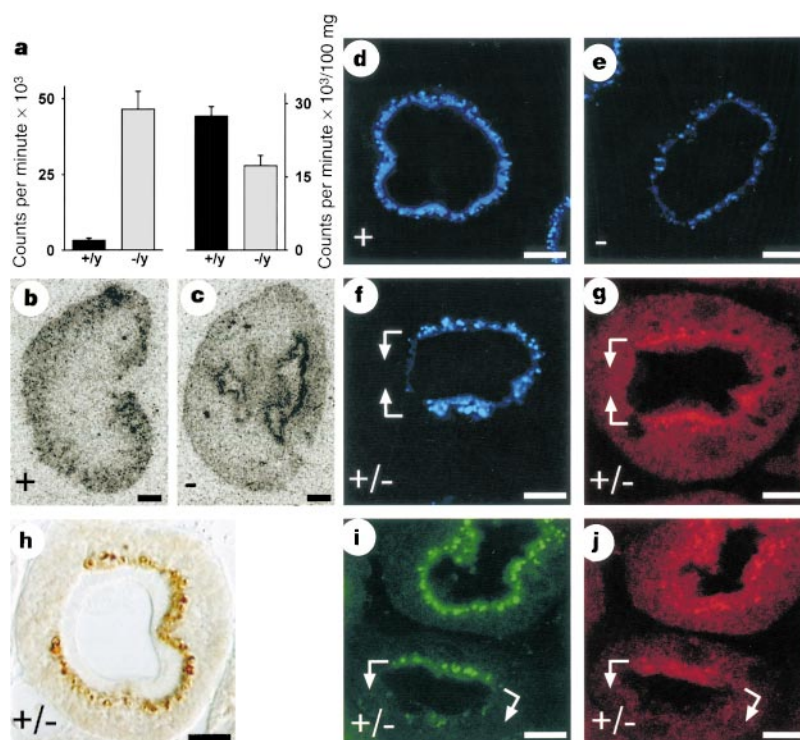


Figure 2 Proximal tubular endocytosis with various *clcn5* genotypes. **a**, ¹²⁵I β₂-microglobulin excreted into the urine (left) and retained in kidney (right). Autoradiographs of kidney sections of WT (**b**) and *clcn5*^{-/-} (**c**) mice fixed 7 min after injection. **d–f**, Uptake of CY5-labelled β-lactoglobulin (blue; 7-min uptake) into PT cells of *clcn5*^{+/-} (**d**) *clcn5*^{-/-} (**e**), and *clcn5*^{+/-} (**f**) animals. The tubule shown in **f** is stained for CIC-5 (red) in **g**. Cells with reduced endocytosis (between arrows) (**f**) lack CIC-5 (**g**). **h**, Uptake of horseradish

peroxidase (HRP) in a *clcn5*^{+/-} female. Cells with reduced endocytosis probably lack CIC-5, because HRP internalization was drastically lower in *clcn5*^{-/-} mice (data not shown). **i**, Uptake of FITC-dextran (green) in a *clcn5*^{+/-} female. CIC-5 protein in the same tubules is shown in **j**. Bars indicate 10 μm, except for **b** and **c** (1 mm). Genotypes are indicated in the lower left corner.

phosphaturia were found in a recent mouse model in which the CIC-5 protein, but surprisingly not its mRNA, was moderately reduced using the transgenic expression of a ribozyme¹⁰.

Low-molecular-weight proteins pass the glomerular filter and are normally reabsorbed and degraded by the proximal tubule (PT). By following the fate of labelled proteins that were injected into the bloodstream, we examined whether the proteinuria of *clcn5*⁻ mice is due to defective tubular endocytosis. Iodinated β_2 -microglobulin, a protein whose urinary concentration is drastically increased in Dent's disease^{1,2,11}, was lost into the urine of *clcn5*⁻ mice in much larger quantities than in WT mice (Fig. 2a). This difference in excretion was mirrored by the amount retained within the kidneys. Autoradiography of kidney sections showed that radioactivity had accumulated in the cortex (largely consisting of PTs) of WT kidneys (Fig. 2b), but that it was found in deeper (probably pelvic) regions of *clcn5*⁻ kidneys (Fig. 2c). This indicates a defect in protein uptake by PT cells.

We injected fluorescently labelled lactoglobulin and fixed the kidneys after seven minutes. WT animals had accumulated substantial amounts of the protein in vesicles below the brush border of PT cells (Fig. 2d). By contrast, *clcn5*⁻ tubules had taken up much less protein (Fig. 2e). CIC-5 is encoded on the X-chromosome and is subjected to random X-chromosomal inactivation in females. *clcn5*^{+/-} females are chimaeras, in which cell patches either express or lack CIC-5 protein (Fig. 2g). In chimaeric animals, cells lacking CIC-5 take up much less protein than neighbouring WT cells (Fig. 2f), showing that CIC-5 influences endocytosis in a cell-autonomous manner. CIC-5 disruption also inhibited the endocytosis of horseradish peroxidase (Fig. 2h) and of FITC-dextran, a non-proteinaceous marker for fluid phase endocytosis (Fig. 2i and j). We also determined total uptake of FITC-dextran into WT and *clcn5*⁻ kidneys by measuring the fluorescence of kidney extracts. The elimination of CIC-5 led to a roughly 70% reduction in dextran uptake.

Proximal tubular cells use megalin as an apical receptor to

endocytose a broad range of luminal proteins¹². These include β_2 -microglobulin¹³, vitamin D binding protein¹⁴ and retinol binding protein¹⁵. CIC-5 expression partially overlapped with megalin in the apical pole of PT cells. The amount of megalin was significantly reduced in cells lacking CIC-5 (Fig. 3a–c), which was confirmed by western analysis (Fig. 3i). CIC-5 co-localizes with the proton pump in the subapical region of the PT and in α -intercalated cells of the distal nephron^{6,16}. In contrast to megalin, *clcn5* disruption did not significantly affect the expression of the H⁺-ATPase in either cell type (Fig. S2 in Supplementary Information).

Several transport proteins in the brush border of PTs are regulated by endocytosis. This includes the Na⁺-phosphate cotransporter NaPi-2 (accounting for more than 70% of brush border phosphate transport¹⁷) and the apical Na⁺/H⁺ exchanger NHE3 (involved in reabsorption of Na⁺, HCO₃⁻ and fluid). Both proteins are partially present in endocytotic compartments^{18–20} and are removed from the plasma membrane in response to PTH^{18,21,22}. The NaPi-2 transporter is then degraded in lysosomes. When mice were kept on normal diet, NaPi-2 antibodies²³ stained intensely the brush border of WT PT cells (Fig. 3d). Some subapical vesicles were labelled as well. The staining in *clcn5*⁻ animals was less intense and the protein showed a different subcellular localization (Fig. 3e and f). The brush border was labelled in the early S1 segment (Fig. 3f), but NaPi-2 resided predominantly in subapical vesicles in most other parts of the PT (Fig. 3e). This was unexpected, because a defect in NaPi-2 endocytosis should instead entail an increased presence in the plasma membrane. In chimaeric *clcn5*^{+/-} females, the brush border was stained as intensely as in the WT, irrespective of the expression of CIC-5 (Fig. 3g and h). Thus, the effect of *clcn5* disruption on steady-state NaPi-2 localization is not cell-autonomous. Western analysis confirmed that the amount of NaPi-2 protein was reduced in *clcn5*⁻, but not in *clcn5*^{+/-} animals (Fig. 3i).

These results indicate that a defective regulation of NaPi-2 plasma membrane localization may cause hyperphosphaturia in *clcn5*⁻ mice

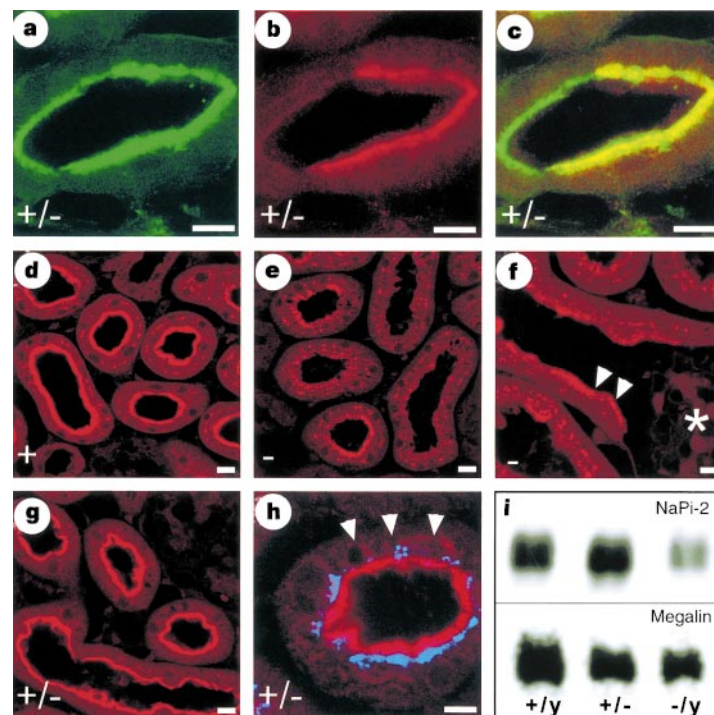


Figure 3 Effect of CIC-5 disruption on transporters and receptors in mice kept on standard diet. A proximal tubule (PT) from a *clcn5*^{+/-} female stained for megalin (green) (a) or for CIC-5 (red) (b). The overlay is shown in c. d–h, NaPi-2 (red) in *clcn5*^{+/-} (d), *clcn5*⁻ (e, f), and *clcn5*^{+/-} tubules (g, h). In all parts (S1–S3) of the PT of *clcn5*^{+/-} (d) or *clcn5*^{+/-} (g, h) mice, NaPi-2 shows predominant plasma membrane staining. In h, endocytosed protein (blue) marks CIC-5 expressing cells. Arrows indicate cells with reduced endocytosis. In

clcn5⁻ tubules (e, f), NaPi-2 staining intensity is reduced. NaPi-2 predominantly localizes to vesicles (e) but is also in the plasma membrane of S1 segments (f, arrowheads) close to glomerula (asterisk). i, Western analysis of kidney membranes from *clcn5*^{+/-}, *clcn5*^{+/-} and *clcn5*⁻ animals using antibodies against NaPi-2 (top) and megalin (bottom). Similar results were obtained in four experiments. Bars indicate 10 μ m. Genotype is indicated at lower left.

and in Dent's disease^{1,24}. Depriving mice of dietary P_i enhances plasma membrane expression of NaPi-2 (ref. 18). Also in *clcn5*⁻ mice, this procedure led to a brush-border expression of NaPi-2, which was now similar to the WT (Fig. 4a and b). When P_i -deprived animals were injected with PTH and analysed after 15 minutes, NaPi-2 was internalized in all PT segments of the WT (Fig. 4c and e), but not of *clcn5*⁻ (Fig. 4d and f) animals. Experiments on *clcn5*^{+/-} mice showed that this effect is cell-autonomous (not shown). Thirty minutes after PTH injection, NaPi-2 began to appear in intracellular vesicles of *clcn5*⁻ tubules but was still predominantly apical (Fig. 4f), whereas it was significantly internalized after one hour (Fig. 4g and h).

The slowed, but otherwise apparently intact, regulation of NaPi-2 by PTH, together with the fact that the changed NaPi-2 localization in *clcn5*⁻ mice under standard diet is not cell-autonomous, suggests that altered PTH levels underlie the vesicular localization of NaPi-2 in knockout mice. Serum PTH was only slightly increased in *clcn5*⁻ mice (see table in Supplementary Information). On the other hand, PTH is also present in the glomerular filtrate. PT cells express PTH receptors at both their basolateral and apical membranes²⁵, and apical PTH receptors are functional²⁶. Furthermore, PTH binds to megalin and is endocytosed and degraded in the PT²⁷. This suggests that the defective endocytosis in *clcn5*⁻ mice and the associated downregulation of megalin (Fig. 3a and c) lead to a progressive increase in luminal PTH levels from the early S1 to late S3 segments. The increased stimulation of luminal PTH receptors will enhance the internalization of NaPi-2 (ref. 26). If this hypothesis is correct, NaPi-2 internalization should not be increased in early S1 segments,

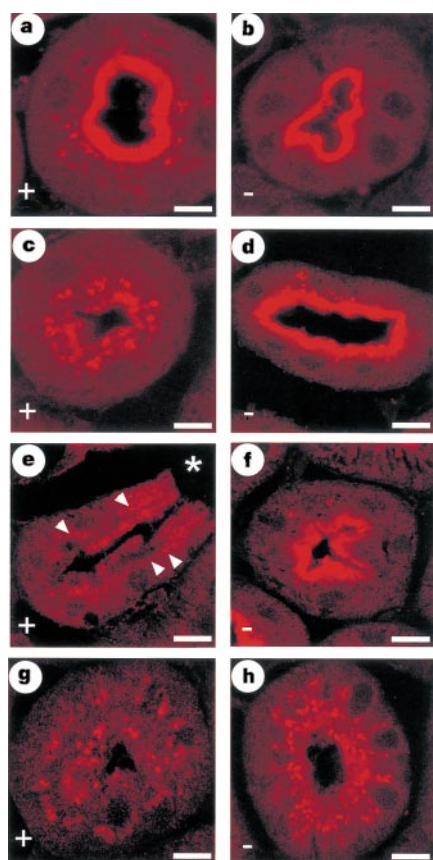


Figure 4 Regulation of NaPi-2 trafficking. **a, b**, NaPi-2 in PTs of phosphate-deprived WT and *clcn5*⁻ mice, respectively. **c–e**, NaPi-2 staining of PTs of phosphate-deprived animals that received PTH for 15 min. **c, e**, *clcn5*⁺; **d, f**, *clcn5*⁻ mice. **e**, An S1 segment leaving the glomerulus (asterisk). Arrowheads indicate that PTH-induced NaPi-2 internalization occurs also in this segment (a control for the differential localization in *clcn5*⁻ mice on standard diet (Fig. 3e and f)). **f**, NaPi-2 in *clcn5*⁻ mice 30 min after PTH. **g, h**, NaPi-2 1 h after PTH administration to phosphate-deprived *clcn5*⁺ (**g**) or *clcn5*⁻ (**h**) mice. Bars indicate 10 μ m.

because PTH levels have not yet risen immediately after filtration. This was indeed observed (Fig. 3f), and contrasts with the uniform internalization of NaPi-2 that is expected for a systemic increase in PTH. This hypothesis additionally predicts an increased urinary loss of PTH. While there is a fourfold increase in urinary PTH in mice lacking megalin²⁷, the less severe defect in megalin-mediated endocytosis in *clcn5*⁻ mice increased urinary PTH by about 1.7-fold (Supplementary Information). Thus, the phosphaturia in *clcn5*⁻ animals may be a consequence of a reduced endocytosis of filtered PTH.

The resulting increase in luminal PTH gives rise to two other predictions. First, the activation of PTH receptors should stimulate the conversion of 25(OH) vitamin D₃ to the active metabolite 1,25(OH)₂-D₃ in the PT. The urinary loss of P_i may also increase 1,25(OH)₂-D₃, as found with the 50% reduction of NaPi-2 gene dosage in *Npt2*^{+/-} mice¹⁷. On the other hand, the reduced endocytosis of vitamin D (VitD) bound to its binding protein leads to a urinary loss of VitD (Supplementary Information) and decreases the availability of the precursor 25(OH) vitamin D₃ within PT cells. Megalin knockout mice, in which receptor-mediated endocytosis of VitD binding protein is totally abolished, display a severe vitamin D deficiency¹⁴. In *clcn5*⁻ mice, serum levels of 25(OH)-D₃ and 1,25(OH)₂-D₃ were reduced about threefold and twofold, respectively (Supplementary Information). Because 1,25(OH)₂-D₃ directly inhibits PTH secretion, this may underlie the slight increase in serum PTH. This contrasts with the slightly elevated levels of 1,25(OH)₂-D₃ and slightly decreased PTH concentrations found in many, but not all, patients with Dent's disease^{1,24}. This difference might be explained by a delicate balance between the opposing effects of the increased stimulation of luminal PTH receptors and the decreased availability of 25(OH) VitD₃, both of which result from reduced apical endocytosis. Depending on species, genetic and nutritional factors, this might lead to either an increase or a decrease in 1,25(OH)₂-D₃ concentration. Because 1,25(OH)₂-D₃ stimulates the intestinal absorption of Ca²⁺ and P_i , its increase may result in hypercalciuria and kidney stones in Dent's disease. Because this is an indirect consequence of the defect in endocytosis, the complexity of the interposed hormonal regulation may explain the lack of nephrocalcinosis in our mouse model and the variable clinical picture of *CLCN5* mutations in humans, where proteinuria and phosphaturia are sometimes the only symptoms^{11,24}.

Second, the apical stimulation of PTH receptors should also promote the steady-state internalization of NHE3 in PTs of *clcn5*⁻ mice kept on normal diet. This is indeed true (Fig. 5a and b). Chimaeric females showed that this effect is not cell-autonomous

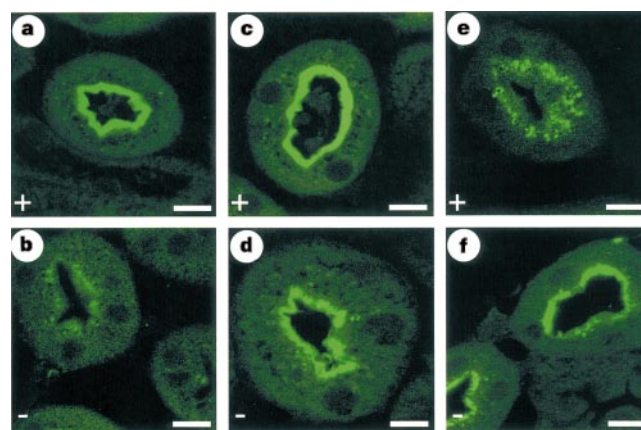


Figure 5 Regulation of NHE3 localization. NHE3 staining (green) of PTs from a WT (**a**) and a *clcn5*⁻ (**b**) mouse kept on standard diet. **c, d**, NHE3 in a phosphate-deprived *clcn5*⁺ (**c**) and *clcn5*⁻ (**d**) mouse. **e, f**, NHE3-stained tubules of a phosphate-deprived WT and *clcn5*⁻ mouse, respectively, 15 min after injection of PTH. Bars indicate 10 μ m. Genotype indicated at lower left.

(not shown), and P_i deprivation led to an apical expression of NHE3 also in knockout mice (Fig. 5c and d). PTH stimulation of P_i -deprived mice revealed a slower rate of NHE3 internalization in cells lacking CIC-5 (Fig. 5e and f). This also shows that the proposed endosomal acidification by NHE3 (refs 20 and 28) cannot compensate for the loss of CIC-5. Because NHE3 is an important determinant of salt and fluid transport in the PT²⁹, the decreased surface expression of NHE3 may contribute to the slightly increased urinary volume of CIC-5 knockout mice.

Thus CIC-5 is crucial for efficient endocytosis in the proximal tubule. CIC-5 is the first intracellular chloride channel for which a role in vesicle trafficking is now firmly established. It is very likely that it provides an electrical shunt for the electrogenic proton pump, thereby allowing the efficient luminal acidification that is required for the proper function of the endocytotic pathway⁶. In preliminary experiments, we have indeed found that endosomes purified from *clcn5* kidneys acidify at slower rates than WT endosomes (N.P. *et al.*, unpublished work). However, we cannot exclude that CIC-5 disruption impairs endocytosis by a mechanism that is unrelated to a change in vesicular pH. CIC-5 disruption causes a dramatic slowing, but not complete elimination, of receptor-mediated endocytosis, fluid-phase endocytosis, and the endocytotic retrieval of plasma membrane transporters. On the other hand, it does not play this role in every tissue. We could not detect a significant defect in endocytosis of asialofetuin in the liver (data not shown), an organ that expresses low levels of CIC-5 (ref. 5 and Fig. 1c). This indicates that other chloride channels perform similar roles in other tissues. Our mouse model strongly suggests that alterations in hormones involved in Ca^{2+} homeostasis, and hyperphosphaturia and hypercalciuria, are indirect effects of defective apical endocytosis of PTH and 25(OH)-D₃; this may explain how a defect in a Cl^- -channel could lead to kidney stones. □

Methods

Generation of *clcn5* mice

Genomic *clcn5* clones were isolated from a mouse λ FixII 129/Sv library (Stratagene) and used to create the targeting vector shown in Fig. 1a. The linearized vector was electroporated into MP12 129/Sv mouse embryonic stem cells. Cells from a correctly targeted clone were injected into C57BL/6J blastocysts that were implanted into foster mothers. Chimaeric males were bred with C57BL/6J females and subsequently crossed into a C57BL/6J background.

In vivo endocytosis and immunocytochemistry

Bovine β-lactoglobulin (Sigma) was labelled with CY5 using a kit (Amersham), and human β₂-microglobulin (Sigma) with ¹²⁵I using Iodogen (Pierce). Labelled proteins, horseradish peroxidase, or FITC-dextran in PBS were infused into the vena cava of anesthetized mice. At the desired time, the kidneys were fixed by perfusion (through the left heart) with 4% paraformaldehyde or periodate lysine paraformaldehyde in phosphate buffered saline (PBS). Tissue samples were embedded in paraffin, and processed as described⁶. Primary antibodies were the rabbit PEP5A1 anti-CIC-5 antibody⁶, a rabbit antibody against NaPi-2 (ref. 23), the mouse monoclonal 1H2 against megalin³⁰, mouse monoclonals against NHE3 (Chemicon), and monoclonal and polyclonal antibodies against the V-type H⁺-ATPase (gifts from S. Gluck and D. Stone). Horseradish peroxidase was revealed using diaminobenzidine, and primary antibodies were detected with secondary antibodies labelled with Alexa 488 or 546 dyes (Molecular Probes). Sections were examined by laser scanning confocal microscopy (Leica).

Membrane trafficking of NaPi-2

For phosphate depletion, mice were kept on a diet having less than 0.04% phosphate (SSNIF, Soest) for six days. To determine the acute effects of PTH, 20 μg rat PTH (Bachem) were injected into the vena cava of anesthetized mice together with CY5-labelled lactoglobulin as a control, and kidneys were fixed as described above. For one-hour experiments, mice were first injected intraperitoneally (i.p.) with 80 μg of PTH, followed after 30 min by an intravenous (i.v.) dose of 20 μg, and kidneys were fixed after a further 30 min.

Serum and urinary analysis

Blood was drawn from the retro-orbital sinus under light ether anaesthesia. Mice were kept in metabolic cages for urine collection. Creatinine, glucose and salts were determined in the department for clinical chemistry. Serum and urinary PTH, serum 1,25(OH)₂-D₃, and calcitonin were determined using radioimmunoassays (Immundiagnostik, Nichols Institute Diagnostica and Immutopics). Urinary samples were directly analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), followed by silver staining or western

blotting using polyclonal antibodies against retinol binding protein (from B. Blaner) or vitamin D binding protein (Dako).

Received 29 June; accepted 29 September 2000.

- Wrong, O. M., Norden, A. G. & Feest, T. G. Dent's disease; a familial proximal renal tubular syndrome with low-molecular-weight proteinuria, hypercalciuria, nephrocalcinosis, metabolic bone disease, progressive renal failure and a marked male predominance. *Q. J. Med.* **87**, 473–493 (1994).
- Scheinman, S. J. X-linked hypercalciuric nephrolithiasis: clinical syndromes and chloride channel mutations. *Kidney Int.* **53**, 3–17 (1998).
- Lloyd, S. E. *et al.* A common molecular basis for three inherited kidney stone diseases. *Nature* **379**, 445–449 (1996).
- Fisher, S. E. *et al.* Isolation and partial characterization of a chloride channel gene which is expressed in kidney and is a candidate for Dent's disease (an X-linked hereditary nephrolithiasis). *Hum. Mol. Genet.* **3**, 2053–2059 (1994).
- Steinmeyer, K., Schwappach, B., Bens, M., Vandewalle, A. & Jentsch, T. J. Cloning and functional expression of rat CLC-5, a chloride channel related to kidney disease. *J. Biol. Chem.* **270**, 31172–31177 (1995).
- Günther, W., Lüchow, A., Cluzeaud, F., Vandewalle, A. & Jentsch, T. J. CLC-5, the chloride channel mutated in Dent's disease, colocalizes with the proton pump in endocytotically active kidney cells. *Proc. Natl Acad. Sci. USA* **95**, 8075–8080 (1998).
- Devuyst, O., Christie, P. T., Courtoy, P. J., Beauwens, R. & Thakker, R. V. Intra-renal and subcellular distribution of the human chloride channel, CLC-5, reveals a pathophysiological basis for Dent's disease. *Hum. Mol. Genet.* **8**, 247–257 (1999).
- Jentsch, T. J., Friedrich, T., Schriever, A. & Yamada, H. The CLC chloride channel family. *Pflügers Arch.* **437**, 783–795 (1999).
- Friedrich, T., Breiderhoff, T. & Jentsch, T. J. Mutational analysis demonstrates that CLC-4 and CLC-5 directly mediate plasma membrane currents. *J. Biol. Chem.* **274**, 896–902 (1999).
- Luyckx, V. A., Ledercq, B., Dowland, L. K. & Yu, A. S. Diet-dependent hypercalciuria in transgenic mice with reduced CLC5 chloride channel expression. *Proc. Natl Acad. Sci. USA* **96**, 12174–12179 (1999).
- Morimoto, T. *et al.* Mutations in CLCN5 chloride channel in Japanese patients with low molecular weight proteinuria. *J. Am. Soc. Nephrol.* **9**, 811–818 (1998).
- Lehste, J. R. *et al.* Megalin knockout mice as an animal model of low molecular weight proteinuria. *Am. J. Pathol.* **155**, 1361–1370 (1999).
- Orlando, R. A. *et al.* Megalin is an endocytic receptor for insulin. *J. Am. Soc. Nephrol.* **9**, 1759–1766 (1998).
- Nykjaer, A. *et al.* An endocytic pathway essential for renal uptake and activation of the steroid 25-(OH) vitamin D₃. *Cell* **96**, 507–515 (1999).
- Christensen, E. I. *et al.* Evidence for an essential role of megalin in transepithelial transport of retinol. *J. Am. Soc. Nephrol.* **10**, 685–695 (1999).
- Sakamoto, H. *et al.* Cellular and subcellular immunolocalization of CLC-5 channel in mouse kidney: colocalization with H⁺-ATPase. *Am. J. Physiol.* **277**, F957–F965 (1999).
- Beck, L. *et al.* Targeted inactivation of Npt2 in mice leads to severe renal phosphate wasting, hypercalciuria, and skeletal abnormalities. *Proc. Natl Acad. Sci. USA* **95**, 5372–5377 (1998).
- Murer, H. *et al.* Posttranscriptional regulation of the proximal tubule NaPi-II transporter in response to PTH and dietary Pi. *Am. J. Physiol.* **277**, F676–F684 (1999).
- Biemesderfer, D. *et al.* Monoclonal antibodies for high-resolution localization of NHE3 in adult and neonatal rat kidney. *Am. J. Physiol.* **273**, F289–F299 (1997).
- D'Souza, S. *et al.* The epithelial sodium-hydrogen antiporter Na⁺/H⁺ exchanger 3 accumulates and is functional in recycling endosomes. *J. Biol. Chem.* **273**, 2035–2043 (1998).
- Traebert, M., Roth, J., Biber, J., Murer, H. & Kaissling, B. Internalization of proximal tubular type II Na-Pi cotransporter by PTH: immunogold electron microscopy. *Am. J. Physiol.* **278**, F148–F154 (2000).
- Zhang, Y. *et al.* In vivo PTH provokes apical NHE3 and NaPi2 redistribution and Na-K-ATPase inhibition. *Am. J. Physiol.* **276**, F711–F719 (1999).
- Custer, M., Lotscher, M., Biber, J., Murer, H. & Kaissling, B. Expression of Na-Pi cotransport in rat kidney: localization by RT-PCR and immunohistochemistry. *Am. J. Physiol.* **266**, F767–F774 (1994).
- Bosio, M., Bianchi, M. L., Lloyd, S. E. & Thakker, R. V. A familial syndrome due to Arg648Stop mutation in the X-linked renal chloride channel gene. *Pediatr. Nephrol.* **13**, 278–283 (1999).
- Kaufmann, M. *et al.* Apical and basolateral parathyroid hormone receptors in rat renal cortical membranes. *Endocrinology* **134**, 1173–1178 (1994).
- Traebert, M., Völkl, H., Biber, J., Murer, H. & Kaissling, B. Luminal and contraluminal action of 1-34 and 3-34 PTH peptides on renal type IIa Na-Pi cotransporter. *Am. J. Physiol.* **278**, F792–F798 (2000).
- Hilpert, J. *et al.* Megalin antagonizes activation of the parathyroid hormone receptor. *J. Biol. Chem.* **274**, 5620–5625 (1999).
- Gekle, M. *et al.* Inhibition of Na⁺-H⁺ exchange impairs receptor-mediated albumin endocytosis in renal proximal tubule-derived epithelial cells from opossum. *J. Physiol. (Lond.)* **520**, 709–721 (1999).
- Schultheis, P. J. *et al.* Renal and intestinal absorptive defects in mice lacking the NHE3 Na⁺/H⁺ exchanger. *Nature Genet.* **19**, 282–285 (1998).
- Zheng, G. *et al.* Organ distribution in rats of two members of the low-density lipoprotein receptor gene family, gp330 and LRP/alpha 2MR, and the receptor-associated protein (RAP). *J. Histochem. Cytochem.* **42**, 531–542 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank G. Boia for technical assistance, P. Gruss for MPI2 ES cells, H. Murer and J. Biber for the NaPi-2 antibody, R. McCluskey for the megalin antibody, S. Gluck and D. K. Stone for proton pump antibodies, B. Blaner for the antibody against retinol binding protein and M. Haddad for chemical analysis of serum and urine samples. This work was supported by grants from the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and the Louis-Jeantet Prize for medicine to T.J.J.

Correspondence and requests for materials should be addressed to T.J.J. (e-mail: jentsch@plexus.uke.uni-hamburg.de).

The Eps8 protein coordinates EGF receptor signalling through Rac and trafficking through Rab5

Letizia Lanzetti*, Vladimir Rybin†, Maria Grazia Malabarba*, Savvas Christoforidis‡, Giorgio Scita*, Marino Zerial† & Pier Paolo Di Fiore*§

* Department of Experimental Oncology, European Institute of Oncology, 20141 Milan, Italy

† Max Planck Institute for Molecular Cell Biology and Genetics, 01307 Dresden, c/o European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany

§ IFOM, The FIRC Institute for Molecular Oncology, 20134 Milan, Italy

‡ Present address: Laboratory of Biological Chemistry, Medical School University of Ioannina, 45110 Ioannina, Greece

How epidermal growth factor receptor (EGFR) signalling is linked to EGFR trafficking is largely unknown. Signalling and trafficking involve small GTPases of the Rho and Rab families, respectively. But it remains unknown whether the signalling relying on these two classes of GTPases is integrated, and, if it is, what molecular machinery is involved. Here we report that the protein Eps8 connects these signalling pathways. Eps8 is a substrate of the EGFR¹, which is held in a complex with Sos1 by the adaptor protein E3b1 (ref. 2), thereby mediating activation of Rac². Through its *src* homology-3 domain, Eps8 interacts with RN-tre³. We show that RN-tre is a Rab5 GTPase-activating protein, whose activity is regulated by the EGFR. By entering in a complex with Eps8, RN-tre acts on Rab5 and inhibits internalization of the EGFR. Furthermore, RN-tre diverts Eps8 from its Rac-activating function, resulting in the attenuation of Rac signalling. Thus, depending on its state of association with E3b1 or RN-tre, Eps8 participates in both EGFR signalling through Rac, and trafficking through Rab5.

RN-tre displays a protein homology domain, which we called TrH³ (for tre homology; also called TBC or PTM⁴), that has been

shown to encode the catalytic core of the GTPase-activating (GAP) activity of Gyp1 and Gyp7, two yeast GAPs for Ypt/Rab GTPases⁵. In RN-tre immunoprecipitates, we detected a Rab5-GAP activity that could be mapped to the TrH-containing region (residues 2–395) of RN-tre (see Supplementary Information). A recombinant TrH domain (glutathione S-transferase (GST)–TrH) stimulated GTP hydrolysis by Rab5, but not by Rab4, Rab11, Rac, Cdc42, Ras or Rho (Fig. 1a). Rab5 displays a higher intrinsic rate of GTPase activity than other Rab proteins (0.12 min^{−1}) (Fig. 1b)⁶. Catalytic amounts of GST–TrH increased the rate of hydrolysis to 0.46 min^{−1} (Fig. 1b), a value in the same range of the GTPase activity measured for Rab5 on the endosome membrane *in vitro*⁷. GST–TrH displayed no effect on nucleotide exchange (see Supplementary Information). There was also a direct interaction between RN-tre and GTP-bound Rab5 (see Supplementary Information).

In a general model for GAP(s) catalysis, two arginine residues are essential to stabilize the transition state of the GTP hydrolysis reaction⁸. In TrH domains⁴, two arginine residues (residues 106 and 150 of RN-tre) are highly conserved. In addition, an aspartate residue (residue 147 of RN-tre) is invariant⁴. We individually mutated Arg 106, Arg 150 and Asp 147 to alanine. Consistent with the proposed mechanism, these mutants failed to display GAP activity on Rab5 (Fig. 1c).

Rab5 regulates trafficking in the early endocytic pathway⁹ and Rab5-GTP is required for the homotypic fusion of early endosomes *in vitro*^{7,10}. Thus, a Rab5-GAP should inhibit early endosome fusion. Indeed, endosome fusion was blocked by GST–TrH, but not by the GAP-defective TrH mutants (Fig. 1d).

Owing to its GAP activity, overexpression of RN-tre should inhibit Rab5-dependent functions *in vivo*, such as endocytosis of the transferrin receptor (TfR)^{9,11}. Overexpression of RN-tre, but not of the GAP-defective R150A mutant, resulted in severe impairment of transferrin internalization (Fig. 2a). Endocytosis of the TfR is a constitutive process, whereas other receptors are internalized upon ligand engagement. The prototype of such receptors is the epidermal growth factor (EGF) receptor (EGFR), whose internalization might also be regulated by Rab5 (ref. 12). Indeed, a dominant-negative mutant of Rab5 (Rab5S34N) inhibited EGF internalization (Fig. 2b). Accordingly, overexpression of RN-tre also inhibited

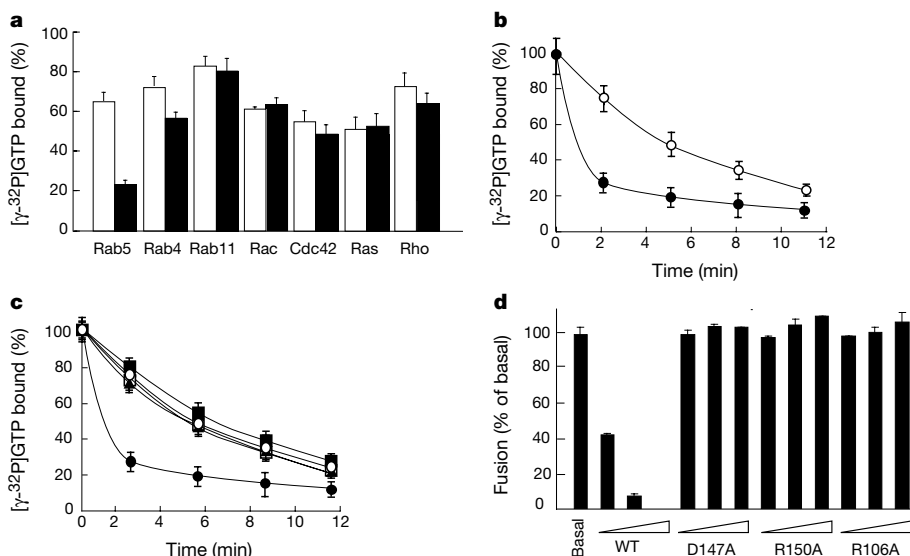


Figure 1 RN-tre is a Rab5-GAP whose catalytic activity is encoded by the TrH domain. **a**, The indicated GTPases (2 μ M) were loaded with [γ -³²P]GTP and incubated with GST (open bars) or purified GST–TrH (100 nM) (filled bars). **b**, Time-dependent kinetic of the GAP activity of GST–TrH (100 nM) (filled circles) or GST (open circles) on [γ -³²P]GTP–Rab5 (2 μ M). **c**, [γ -³²P]GTP–Rab5 (2 μ M) was incubated with either wild-type GST–TrH (filled circles) or GST–TrH(R106A) (filled squares), or GST–TrH(R150A) (filled triangles), or

GST–TrH(D147A) (open squares) (100 nM) or with GST (open circles). **d**, Early endosome fusion assay. The assay was performed in the presence of HeLa cells cytosol and ATP-regenerating mix (energy) (basal = 100% fusion) or cytosol and energy plus increasing concentrations (7, 35 and 178 nM) of GST–TrH wild-type (WT) or the three GST–TrH mutants (D147A, R150A, R106A).

EGFR endocytosis in a GAP-dependent manner (Fig. 2b). We also showed that EGFR-dependent signalling attenuates the GAP activity of RN-tre. After exposure of intact cells to EGF, we recovered significantly decreased levels of Rab5-GAP activity in RN-tre

immunoprecipitates, with respect to unstimulated cells (Fig. 2c). RN-tre is phosphorylated on serine after EGF stimulation within a time-frame similar to that of the downregulation of its GAP activity (data not shown); it will be therefore of interest to determine

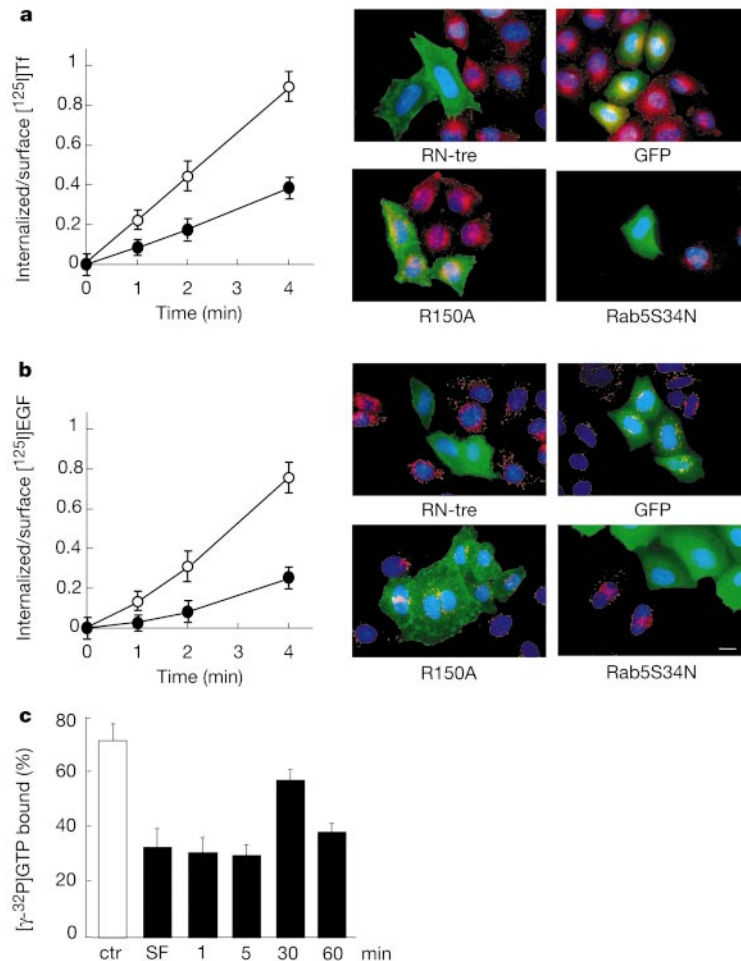


Figure 2 RN-tre is a Rab5-GAP *in vivo* and its activity is negatively regulated by EGFR. **a, b**, Internalization of Tf (**a**) and EGF (**b**). Left, HeLa cells were infected with the RN-tre retrovirus (filled circles) or a control empty retrovirus (open circles) (see also Supplementary Information). Mass populations were tested for internalization of [¹²⁵I]Tf (1 μg ml⁻¹; **a**) or [¹²⁵I]-EGF (2 ng ml⁻¹; **b**). Right, HeLa cells were transiently transfected with GFP-RN-tre, GFP alone, the mutant GFP-RN-tre(R150A) or the dominant-negative

Rab5 mutant (Rab5S34N) and exposed to rhodamine-Tf (**a**) or rhodamine-EGF (**b**). Red, rhodamine-transferrin or EGF; green, GFP-tagged proteins; blue, DAPI counterstain. Scale bar, 10 μm. **c**, [γ-³²P]GTP-Rab5 was incubated with anti-HA immunoprecipitates from Cos-7 cells mock transfected (open bar) or transfected with HA-RN-tre (filled bars). Where indicated, cells were serum starved (SF) and then treated with EGF (100 ng ml⁻¹ at 37 °C) for the indicated lengths of time, before lysis (see also Supplementary Information).

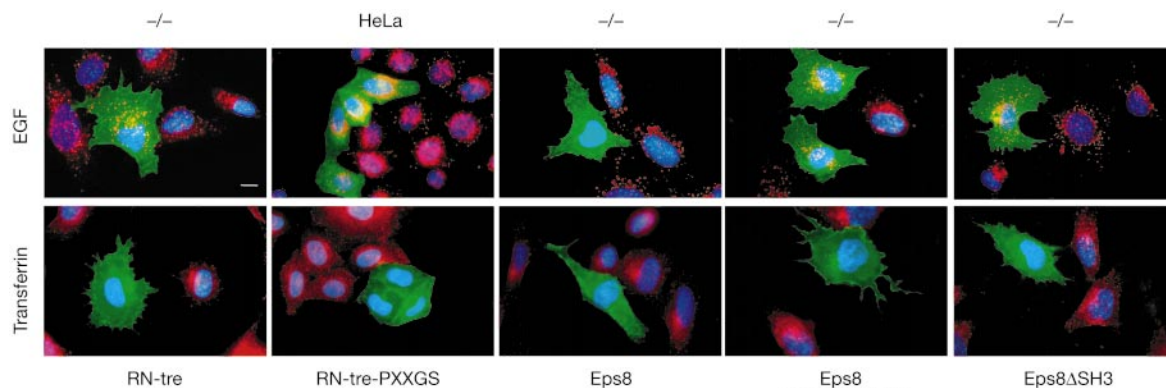


Figure 3 Binding to Eps8 regulates RN-tre *in vivo*. Ep8^{-/-} fibroblasts or HeLa cells (as indicated on top) were transfected with the indicated plasmids (bottom) and tested for their ability to internalize rhodaminated EGF or transferrin. The RN-tre-PXXGS mutant bears a mutation (DY→GS) that renders it unable to interact with the SH3 domain of Eps8

(see Supplementary Information). The Ep8ΔSH3 mutant¹³ bears a 5-amino-acid deletion in its SH3 domain that renders it unable to interact with RN-tre (see Supplementary Information). Results are typical and representative of three independent experiments. Scale bar, 10 μm.

whether this post-translational modification affects the enzymatic activity of RN-tre. Whatever the case, the sum of our results establishes RN-tre as a GAP that modulates the activity of Rab5 both *in vitro* and *in vivo*, and whose activity is controlled by EGFR.

RN-tre binds to the *src* homology (SH)-3 of Eps8 (ref. 3), an activator of Rac-specific guanine nucleotide exchange factor(s) (GEF)². We therefore tested whether Rab5 and Rac signalling might be co-regulated. In Eps8 null fibroblasts (Eps8^{-/-}), overexpression of RN-tre failed to inhibit EGFR internalization, while still inhibiting TfR endocytosis (Fig. 3). In Fig. 3, we further showed (1) that in HeLa cells a mutant RN-tre (RN-tre-PXXGS), which cannot bind to Eps8 (Supplementary Information), failed to inhibit EGFR endocytosis; (2) that transfection of Eps8 in Eps8^{-/-} fibroblasts restored the ability of RN-tre, but not of RN-tre-PXXGS, to inhibit EGFR internalization; (3) that transfection of Eps8^{-/-} fibroblasts with a mutant Eps8, carrying a 5-amino-acid deletion in its SH3 domain (Eps8ΔSH3; ref. 13) and no longer able to bind to RN-tre (see Supplementary Information), failed to restore the effect of RN-tre on EGFR internalization; (4) that TfR internalization was still inhibited, by RN-tre or RN-tre-PXXGS, in all of the above conditions. Thus, the interaction between RN-tre and Eps8 is necessary for the regulation of ligand-dependent but not of constitutive endocytosis.

Eps8 activates Rac by stimulating the Rac-GEF activity of Sos-1 (ref. 2). *In vivo*, a tri-complex exists in which a scaffolding protein, E3b1, brings together Eps8 and Sos-1. All three components are needed for the complex to function as a Rac-GEF². Eps8 interacts with both E3b1 and RN-tre through its SH3 (ref. 14). Thus, Eps8–E3b1 and Eps8–RN-tre complexes should be mutually exclusive.

Indeed, the overexpression of RN-tre decreased the co-immunoprecipitation between Eps8 and E3b1 (Fig. 4a), and resulted in the attenuation of Rac-specific signalling. Using activation of c-Jun aminoterminal kinase (JNK) as a readout, we showed that overexpression of RN-tre in Cos-7 cells resulted in a threefold decrease in EGF-induced JNK activity, but left mitogen-activated protein kinase (MAPK) activity unperturbed (Fig. 4b). Expression of the RN-tre(R150A) (R150A) mutant also inhibited JNK (Fig. 4b), suggesting that the physical interaction of RN-tre with Eps8, rather than the GAP activity, was responsible for the effect.

We validated this hypothesis in Eps8^{-/-} fibroblasts, in which signalling from Ras to Rac, but not from Ras to Raf, is impaired². Accordingly, a threefold reduction of the EGF-dependent JNK activity, but not MAPK, was observed in Eps8^{-/-} fibroblasts, as compared with wild-type cells (Fig. 4c). The persistence of some EGF-dependent JNK induction in Eps8^{-/-} cells might reflect the existence of Rac-independent pathways that also contribute to the activation of this enzyme by growth factors¹⁵. As expected, re-expression of Eps8 in Eps8^{-/-} cells resulted in the restoration of full EGF-dependent JNK activity (Fig. 4d). Overexpression of RN-tre in Eps8^{-/-} fibroblasts did not modify the residual levels of EGF-induced JNK activity (Fig. 4d). However, concomitant expression of Eps8 and RN-tre in Eps8^{-/-} cells abrogated the Eps8-induced effect on JNK activity, which was indistinguishable from that of untransfected Eps8^{-/-} cells (Fig. 4d).

We have uncovered an unexpected link between the machinery of receptor trafficking and signalling, in that two alternative complexes, Eps8–E3b1 and Eps8–RN-tre control Rac and Rab5 signalling, respectively. Eps8, which is partly localized at the plasma

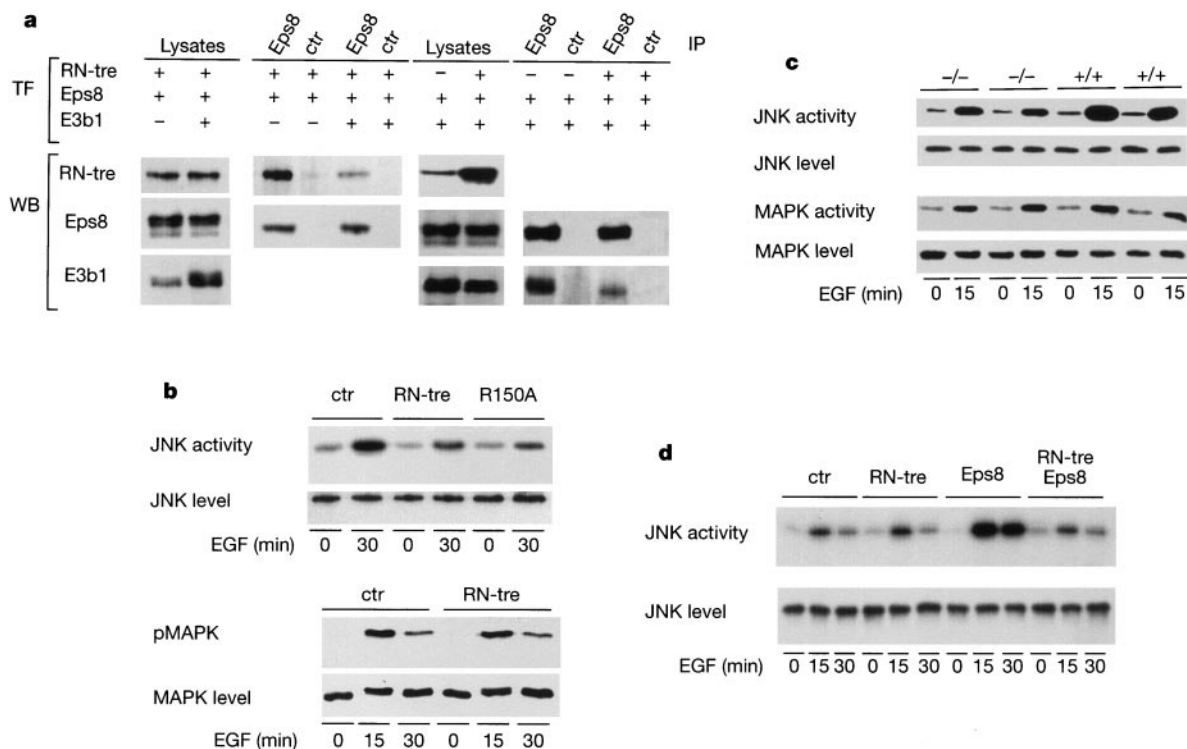


Figure 4 Binding of RN-tre to Eps8 results in the attenuation of Rac-mediated signalling. **a**, Cos-7 cells were transfected (TF) with Eps8, RN-tre and E3b1, as indicated. Cell lysates (1 mg) were immunoprecipitated (IP) with anti-eps8 or irrelevant (ctr) sera and blotted (WB) with the indicated sera. **b**, Cos-7 cells were transfected with HA–JNK or HA–MAPK alone (ctr), or together with RN-tre or the RN-tre(R150A) mutant (see also Supplementary Information). JNK activity was measured by an immunocomplex kinase assay on anti-HA immunoprecipitates, using c-Jun(79) as a substrate. MAPK activity was measured by immunoblotting anti-HA immunoprecipitates with anti-phosphoMAPK (New England Biolabs). Aliquots of the reactions were also immunoblotted with anti-HA to show equal

loading. **c**, Eps8^{-/-} and Eps8^{+/+} clones (2 clones each², indicated on top) were analysed for JNK and MAPK activities by immunocomplex kinase assays on anti-JNK-1 and anti-MAPK immunoprecipitates, using c-jun(79) and MBP as substrates. Aliquots of the reactions were also immunoblotted with specific anti-JNK and anti-ERK1 antibodies to show equal loading. **d**, Eps8^{-/-} cells were transfected with the plasmids indicated on the top, together with HA–JNK and tested for JNK activity as in **b** (see also Supplementary Information). The levels of JNK are also shown. Results are typical and representative of three independent experiments.

membrane¹⁶ and physically interacts with the EGFR¹⁷, is likely to recruit and coordinate the activity of the two types of signalling complexes. We have established that RN-tre is a Rab5 GAP. By acting on Rab5, RN-tre regulates both the constitutive and the regulated (for example, EGF-dependent) receptor internalization. The latter activity strictly depends on the interaction of RN-tre with Eps8. Recruitment of RN-tre would decrease Rab5 activity, thereby inhibiting receptor internalization and, consequently, prolonging receptor signalling at the plasma membrane. Alternatively, recruitment of E3b1 (and Sos1) couples the receptor with downstream Rac activation². Possibly, both Eps8–RN-tre and Eps8–E3b1 complexes are formed simultaneously and concur to enhance receptor signalling. Or maybe the formation of the two complexes is spatially and/or temporally uncoupled, which might happen under rate-limiting concentrations of Eps8 in a restricted signalling microenvironment, such as caveolae¹⁸. In this last case, competition between RN-tre and E3B1 for Eps8 might result in more complex outcomes, dependent on the stoichiometry and temporal hierarchy of the two interactions. Whatever the case, our results establish a new modality of integration of signals emanating from Rho-like and Rab-like GTPases: an unexpected connection that warrants further investigation. □

Methods

Expression vectors and antibodies

We engineered haemagglutinin (HA)-, green fluorescent protein (GFP)- or GST-tagged full-length RN-tre, truncations and point mutants obtaining appropriate fragments by recombinant PCR, or by endonuclease digestion, followed by subcloning in the pcDNAHA1 (ref. 19), pEGFP-C1 (Clontech), or pGEX-4T2 (Pharmacia) vectors.

We engineered the RN-tre retrovirus by cloning the untagged full-length RN-tre in the pBMN retroviral vector, under the control of the CMV promoter. This vector also drives the transcription of the GFP gene from a LTR promoter. After rescue of infectious retroviruses, infection of HeLa cells was performed under conditions yielding more than 80% of infected cells, as assessed by expression of GFP.

We obtained recombinant small GTPases of the Rab family by subcloning the appropriate recombinant PCR fragments into pGEX vectors. All constructs were verified by DNA sequencing. Further details are available upon request. Antibodies used were an anti-HA monoclonal antibody (12CA5, BABCO) and affinity purified anti-RN-tre³, anti-E3b1 (ref. 20), and anti-Eps8 (ref. 1) polyclonal sera. We carried out immunoprecipitation and immunoblotting as described¹.

Biochemical and functional assays

For GAP assays on immunoprecipitates, 6 mg of total cellular lysates were immunoprecipitated with anti-HA and subsequently incubated for 5 min at 20 °C in the presence of [γ -³²P]GTP-loaded²¹ GTPases (0.15 μ M). Incubation mixtures were subjected to filter-binding assays as described²¹. We carried out GAP assays with GST–TrH as described²¹, using 2 μ M of substrate [γ -³²P]GTP-loaded Rab5 and catalytic concentrations of GST–TrH (100 nM) at 30 °C for 2 min, unless otherwise indicated. GAP activity was expressed as the percentage of non-hydrolysed [γ -³²P]GTP that remained bound to the filters, relative to the radioactivity at time 0. Values are the mean $\pm$ s.e. of three independent experiments performed in triplicate.

We carried out endosome fusion assays²² and MAPK and JNK assays¹⁹ as described. Internalization assays, using either fluorochrome-conjugated or radiolabelled ligands have also been described^{23,24}. In the case of fluorochrome-conjugated ligands, exposure to the ligands was for 15 min.

Received 9 June; accepted 23 August 2000.

1. Fazioli, F. *et al.* Eps8, a substrate for the epidermal growth factor receptor kinase, enhances EGF-dependent mitogenic signals. *EMBO J.* **12**, 3799–3808 (1993).
2. Scita, G. *et al.* Eps8 and E3B1 transduce signals from Ras to Rac. *Nature* **401**, 290–293 (1999).
3. Matoskova, B., Wong, W. T., Nomura, N., Robbins, K. C. & Di Fiore, P. P. RN-tre specifically binds to the SH3 domain of eps8 with high affinity and confers growth advantage to NIH3T3 upon carboxy-terminal truncation. *Oncogene* **12**, 2679–2688 (1996).
4. Neuwald, A. F. A shared domain between a spindle assembly checkpoint protein and Ypt/Rab-specific GTPase-activators. *Trends Biochem. Sci.* **22**, 243–244 (1997).
5. Albert, S., Will, E. & Gallwitz, D. Identification of the catalytic domains and their functionally critical arginine residues of two yeast GTPase-activating proteins specific for Ypt/Rab transport GTPases. *EMBO J.* **18**, 5216–5225 (1999).
6. Simon, I., Zerial, M. & Goody, R. S. Kinetics of interaction of Rab5 and Rab7 with nucleotides and magnesium ions. *J. Biol. Chem.* **271**, 20470–20478 (1996).
7. Rybin, V. *et al.* GTPase activity of Rab5 acts as a timer for endocytic membrane fusion. *Nature* **383**, 266–269 (1996).
8. Scheffzek, K., Ahmadian, M. R. & Wittinghofer, A. GTPase-activating proteins: helping hands to complement an active site. *Trends Biochem. Sci.* **23**, 257–262 (1998).
9. Bucci, C. *et al.* The small GTPase rab5 functions as a regulatory factor in the early endocytic pathway. *Cell* **70**, 715–728 (1992).

10. Gorvel, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. rab5 controls early endosome fusion in vitro. *Cell* **64**, 915–925 (1991).
11. McLauchlan, H. *et al.* A novel role for Rab5-GDI in ligand sequestration into clathrin-coated pits. *Curr. Biol.* **8**, 34–45 (1998).
12. Papini, E. *et al.* The small GTP binding protein rab7 is essential for cellular vacuolation induced by *Helicobacter pylori* cytotoxin. *EMBO J.* **16**, 15–24 (1997).
13. Kishan, K. V., Scita, G., Wong, W. T., Di Fiore, P. P. & Newcomer, M. E. The SH3 domain of Eps8 exists as a novel intertwined dimer. *Nature Struct. Biol.* **4**, 739–743 (1997).
14. Mongioli, A. M. *et al.* A novel peptide–SH3 interaction. *EMBO J.* **18**, 5300–5309 (1999).
15. Russell, M., Lange-Carter, C. A. & Johnson, G. L. Direct interaction between Ras and the kinase domain of mitogen-activated protein kinase kinase kinase (MEKK1). *J. Biol. Chem.* **270**, 11757–11760 (1995).
16. Provenzano, C. *et al.* Eps8, a tyrosine kinase substrate, is recruited to the cell cortex and dynamic F-actin upon cytoskeleton remodeling. *Exp. Cell Res.* **242**, 186–200 (1998).
17. Castagnino, P. *et al.* Direct binding of eps8 to the juxtamembrane domain of EGFR is phosphotyrosine- and SH2-independent. *Oncogene* **10**, 723–729 (1995).
18. Anderson, R. G. The caveolae membrane system. *Annu. Rev. Biochem.* **67**, 199–225 (1998).
19. Coso, O. A. *et al.* The small GTP-binding proteins Rac1 and Cdc42 regulate the activity of the JNK/SAPK signalling pathway. *Cell* **81**, 1137–1146 (1995).
20. Biesova, Z., Piccoli, C. & Wong, W. T. Isolation and characterization of e3B1, an eps8 binding protein that regulates cell growth. *Oncogene* **14**, 233–241 (1997).
21. Ridley, A. J. *et al.* rho family GTPase activating proteins p190, bcr and rhoGAP show distinct specificities in vitro and in vivo. *EMBO J.* **12**, 5151–5160 (1993).
22. Horiuchi, H. *et al.* A novel Rab5 GDP/GTP exchange factor complexed to Rabaptin-5 links nucleotide exchange to effector recruitment and function. *Cell* **90**, 1149–1159 (1997).
23. Carbone, R. *et al.* eps15 and eps15R are essential components of the endocytic pathway. *Cancer Res.* **57**, 5498–5504 (1997).
24. Sorkin, A., Di Fiore, P. P. & Carpenter, G. The carboxyl terminus of epidermal growth factor receptor/erbB-2 chimerae is internalisation impaired. *Oncogene* **8**, 3021–3028 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank B. Matoskova for reagents and for helpful discussions, and N. Gholami-Shangolabad for technical assistance. This work was supported by grants from Associazione Italiana Ricerca sul Cancro, from the Armenise-Harvard Foundation and from the CNR (Progetto Biotecnologie).

Correspondence and requests for materials should be addressed to P.P.D.F. (e-mail: pdifiore@ieo.it).

Deacetylation of p53 modulates its effect on cell growth and apoptosis

Jianyuan Luo*, Fei Su*, Delin Chen, Ariel Shiloh & Wei Gu

Institute of Cancer Genetics, and Department of Pathology, College of Physicians & Surgeons, Columbia University, 1150 St. Nicholas Avenue, New York, New York 10032, USA

* These two authors contributed equally to this work

The p53 tumour suppressor is a transcriptional factor whose activity is modulated by protein stability and post-translational modifications including acetylation^{1–4}. The mechanism by which acetylated p53 is maintained *in vivo* remains unclear. Here we show that the deacetylation of p53 is mediated by an histone deacetylase-1 (HDAC1)-containing complex. We have also purified a p53 target protein in the deacetylase complexes (designated PID; but identical to metastasis-associated protein 2 (MTA2)), which has been identified as a component of the NuRD complex^{5–7}. PID specifically interacts with p53 both *in vitro* and *in vivo*, and its expression reduces significantly the steady-state levels of acetylated p53. PID expression strongly represses p53-dependent transcriptional activation, and, notably, it modulates p53-mediated cell growth arrest and apoptosis. These results show that deacetylation and functional interactions by the PID/MTA2-associated NuRD complex may represent an important pathway to regulate p53 function.

Previous studies have indicated that p300/CBP strongly potentiates p53-dependent transcriptional activation, and that acetylation

of p53 by p300 markedly stimulates its sequence-specific DNA-binding activity^{8–11}. However, the proportion of acetylated p53 is enhanced when cells are treated with deacetylase inhibitors such as Trichostatin A (TSA) (ref. 8), implying a tight regulation of p53 acetylation by deacetylases *in vivo*. To test the hypothesis that p53 interacts physically with one of many deacetylase complexes, we used a glutathione *S*-transferase (GST)–p53 affinity column to search for components of deacetylase complexes that may interact with p53. Western blot analysis showed that the eluted GST–p53 associated proteins contained human histone deacetylase HDAC1 (Fig. 1a). However, we also found that p53 failed to bind to highly purified baculovirus-produced recombinant HDAC1 (Fig. 1b), indicating an indirect interaction between p53 and HDAC1.

Studies on mammalian histone deacetylase complexes have indicated that HDAC1 is present in several deacetylase complexes that contain a number of polypeptides^{5–7,12–16}. We proposed that p53 may interact with an HDAC1-associated deacetylase complex through an unknown target protein, and applied a two-step strategy to identify its target protein. First, we purified total HDAC1-containing protein complexes from cell extracts, and second, we isolated the p53 target protein from these complexes.

To obtain purified HDAC1-containing protein complexes, we created an HeLa-derived cell line that stably expresses Flag-epitope-tagged HDAC1. This strategy has been used to purify large amounts of protein complexes from human cells^{17,18}. To isolate Flag–HDAC1 associated complexes, nuclear extracts from the stably transfected HDAC-1 cell line were separated by affinity chromatography and bound proteins were eluted with the Flag peptide. To identify the p53 target protein from these complexes, we performed an affinity-binding assay on the GST–p53 affinity column with dissociated protein complexes (see Methods). There were a number of proteins in HDAC1-associated complexes (Fig. 1c); however, only one protein (relative molecular mass ~75,000 ($M_r \approx 75K$)) bound to the GST–p53 column after dissociation of the HDAC1 complex (Fig. 1c, lane 4). We named this protein PID, for p53 target protein in the deacetylase complexes. After a large preparation, we obtained enough material of the p75 band for peptide sequencing, and determined a total of three peptide sequences (see Supplementary

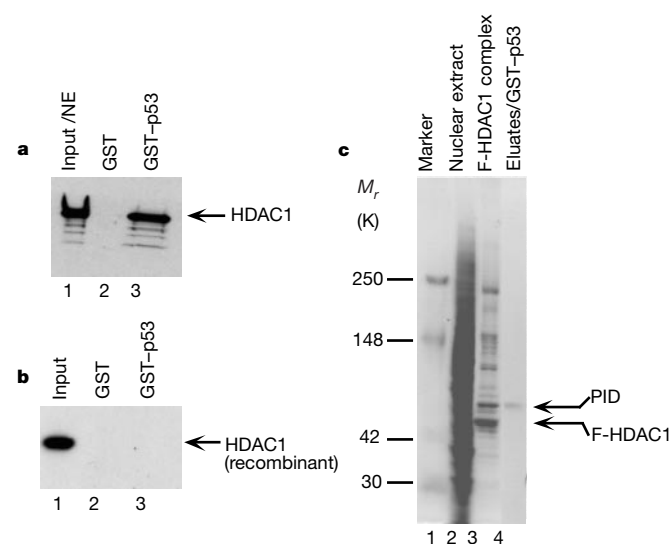


Figure 1 Indirect interactions between p53 and HDAC1, and purification of PID. **a**, Western blot analysis of input HeLa nuclear extract (lane 1), the eluate from the indicated GST (lane 2), or GST–p53 (lane 3) affinity columns with anti-HDAC1. **b**, Western blot analysis of input highly purified recombinant HDAC1 (lane 1) and eluates from either GST (lane 2) or GST–p53 (lane 3) with anti-HDAC1. **c**, Colloidal blue staining of an SDS–PAGE gel containing protein marker (lane 1), HeLa nuclear extract (lane 2), the highly purified HDAC1-associated complexes (F-HDAC1 complex) (lane 3) and the p53-interacting protein eluted from the GST–p53 column (lane 4).

Information). Significantly, however, PID is identical to the human MTA2 polypeptide^{5–7} and homologous to MTA1 (refs 19–21), both of which are present in respective nucleosome remodelling and histone deacetylation (NuRD) complexes that are involved in both nucleosome remodelling and histone deacetylation^{5–7,12–14}. Thus, our data indicates that p53 can be targeted to the NuRD complex through direct interaction with PID/MTA2.

To determine whether p53 directly interacts with PID *per se*, we carried out *in vitro* binding assays with bacterially produced PID and GST–p53 proteins (Fig. 2a). After a short incubation at 4 °C, PID was strongly bound to immobilized GST–p53 but not to immobilized GST alone (Fig. 2a, lanes 6 versus 7). Moreover, PID specifically bound the amino-terminal domain of p53 (Fig. 2a, lane

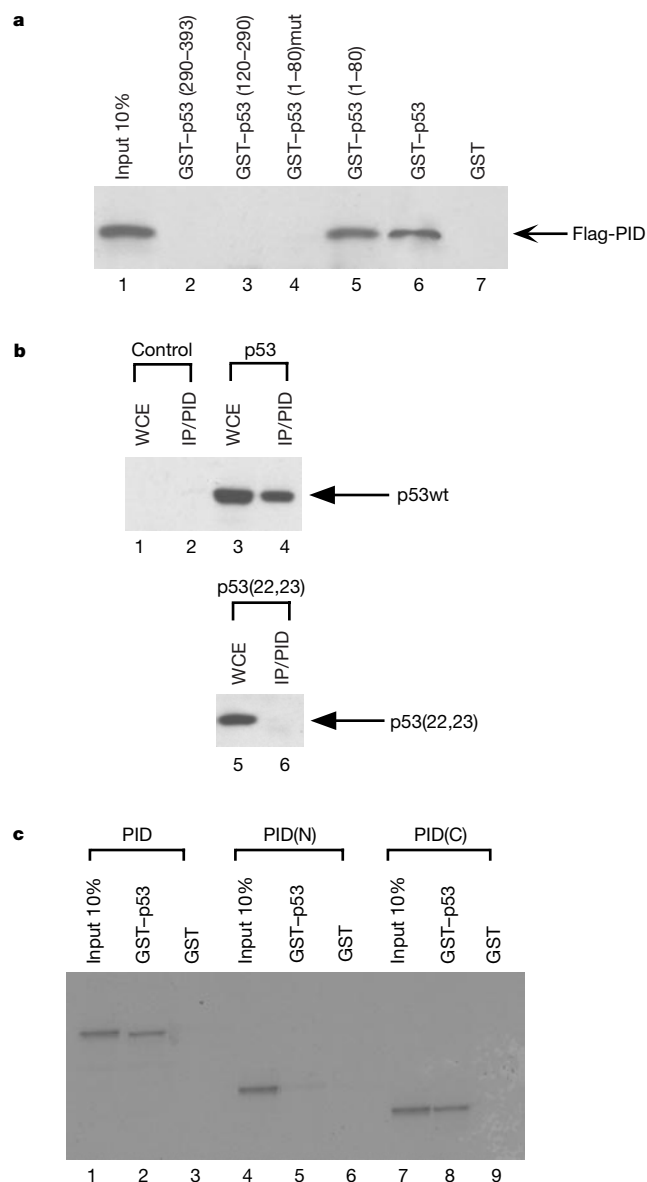


Figure 2 PID interacts with p53 both *in vitro* and *in vivo*. **a**, Direct interactions of recombinant Flag–PID with GST–p53. Western blot analysis of the input and eluates from indicated GST columns by anti-PID antibody (lanes 2–7). **b**, PID interacts with p53 in SaoS-2 cells. Western blot analyses of the indicated whole-cell extract (WCE) and the PID immunoprecipitates (IP/PID) prepared from the transfected SaoS-2 cells with anti full-length p53 polyclonal antibody. **c**, p53 interacts with the N-terminus of PID. The GST–p53 full-length protein was used in GST–pull down assays with different *in vitro* translated ³⁵S-labelled full-length PID protein (PID) (lanes 1–3), the C-terminal domain (1–370) (PID(N)) (lanes 4–6), or the C-terminal domain (371–668) (PID(C)) (lanes 7–9).

5), but not the central DNA-binding domain or the carboxy-terminal domain of p53 (Fig. 2a, lanes 2 and 3). Notably, a double point mutation in the p53 transactivation domain (GST-p53(1–80)mut) (refs 22, 23) showed no detectable binding to PID (Fig. 2a, lane 4). We tested for *in vivo* interactions between p53 and PID by immunoprecipitating extracts from transiently transfected SaoS-2 cells with anti-Flag monoclonal antibody M2. p53 was detected in the immunoprecipitate obtained from SaoS-2 cells co-transfected with constructs encoding Flag-PID and wild-type p53 (Fig. 2b, lanes 3 and 4), but not in the immunoprecipitates from cells transfected with the Flag-PID construct alone (Fig. 2b, lanes 1 and 2). In contrast, a p53 mutant protein (p53(22,23)) containing a double point mutation in the N-terminal domain showed no detectable binding to PID in the same assay (Fig. 2b, lanes 5 and 6), indicating that the binding between p53 and PID is highly specific.

We next established a SaoS-2 derived cell line that stably expresses Flag-PID to obtain the highly purified PID-associated deacetylase complex for functional analysis. Nuclear extracts from these cells were subjected to affinity chromatography on M2-agarose and the Flag-PID associated deacetylase complex eluted with the Flag peptide. Western blot analysis indicated that this complex contains both Flag-PID and HDAC1, but not Sin3 (Fig. 3a, lane 2). Further analyses also indicated that the chromatin remodelling protein CHD4, the tumour-suppressor Rb-interacting protein RbAp48, and the methyl-CpG binding protein MBD3 are also present in the complex (Fig. 3a, lane 2). These data indicate that the Flag-PID complex purified from SaoS-2 cell is an NuRD complex^{5–7,12–14}. This complex was then tested for deacetylation activity on p53 *in vitro*. As indicated in Fig. 3b, ¹⁴C-labelled acetylated p53 was deacetylated by

the purified PID-associated complex (lane 2), but not by the control eluate (lane 3). In addition, the deacetylase inhibitor TSA significantly abrogated the p53 deacetylase activity of this complex (Fig. 3b, lane 4). To establish the role of PID in p53 deacetylation *in vivo*, we used an acetylated p53-specific antibody (see Methods) to monitor the steady-state levels of acetylated p53 *in vivo*. A high level of acetylated p53 was found in the cells co-transfected with p300 and p53 (Fig. 3c, lanes 5 versus 4); however, p53 acetylation was significantly inhibited by PID expression (Fig. 3c lanes 6 versus 5). Notably, the total amount of acetylated p53 was elevated when the cells were treated with TSA (Fig. 3c, lanes 2 versus 5); however, the level of p53 acetylation in TSA-treated cells was not affected by PID expression (Fig. 3c, lanes 2 versus 3), indicating that PID-associated HDAC1 deacetylase activity is required for PID-mediated deacetylation of p53.

To test the effect of PID expression on p53-mediated transcriptional activation, we co-transfected mouse embryo fibroblasts (MEFs) (p53^{−/−}) with the p53 and PID expression vectors, along with a reporter construct containing the minimal p21 promoter. PID repressed p53-mediated transactivation in a dose-dependent manner (Fig. 4a), but had no effect on the transactivation function of Gal-VP16 (Fig. 4b). Also, the PID(N) mutant, which fails to bind p53 (Fig. 2c), showed no effect (Fig. 4c), indicating that the physical interaction with p53 is crucial for PID-dependent repression. Significantly, the p53(K-R) mutant, which was constructed by replacing arginine residues at Lys 370, Lys 372, Lys 373, Lys 381 and Lys 382 (ref. 10), was partially impaired in the PID-mediated repression (Fig. 4d), suggesting that PID-mediated repression of p53 transactivation functions acts, in part, through deacetylation of

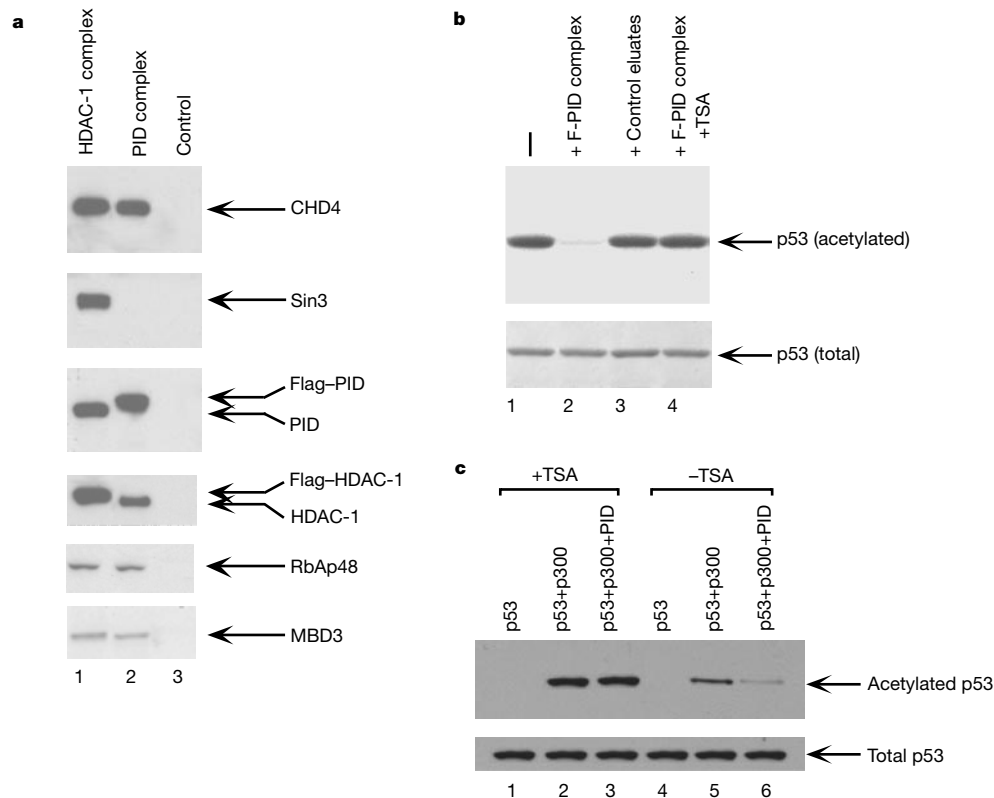


Figure 3 PID is involved in p53 deacetylation. **a**, Western blot analyses of the purified Flag-HDAC1 complexes (lane 1), the purified Flag-PID complex (lane 2) and the control eluate (lane 3) derived from untransfected SaoS-2 cells with anti-CHD4, anti-Sin3, anti-PID, anti-HDAC1, anti-RbAp48 and anti-MBD3. **b**, Deacetylation of p53 by purified Flag-PID complex. The *in vitro* acetylated p53 (lane 1) was incubated with the control eluate (lane 3), the purified PID-associated complex (lane 2), or the complex in the presence of 10 μ M (lane 4) deacetylase inhibitor TSA for 30 min at 30 °C. The proteins were analysed

by resolution on SDS-PAGE and autoradiography (top) or Coomassie blue staining (bottom). **c**, Reduction of the steady-state levels of acetylated p53 by PID expression. Western blot analysis of whole-cell extracts from the H1299 cells transfected with p53 (lanes 1, 4), co-transfected with p53 and p300 (lanes 2, 5), or co-transfected with p53, p300 and PID (lanes 3, 6) by acetylated p53-specific antibody (top) or DO-1 (bottom) for total p53. Cells were either untreated (lanes 4–6) or treated with 500 nM TSA (lanes 1–3).

p53 at the C terminus.

We tested the biological role of PID by examining its effects on known cellular functions of p53 using both colony formation and apoptosis assays. The results revealed that although p53 strongly inhibits SaoS-2 cell growth, the p53-mediated growth inhibition was effectively abrogated by co-transfection with PID (Fig. 5a). This result was corroborated by western blot analysis in which we investigated whether PID expression affected the p53-mediated activation of endogenous p21 gene expression. As indicated in Fig. 5b, p53 strongly activated endogenous p21 gene expression (lanes 2 versus 1); however, the effect was significantly impaired by PID expression (lanes 3 versus 2). Finally, the effect of PID on p53-mediated apoptosis in cells was followed by flow cytometric analysis of transiently transfected cells. H1299 cells were transfected with p53 alone or co-transfected with p53 and PID. Overexpression of p53 alone induced significant apoptosis (29.3% sub-G1) (Fig. 5c). However, co-transfection of p53 with PID significantly reduced the level of apoptosis (12.4% sub-G1), whereas the mutant PID(N) had no effect. Similar results were also obtained by using an annexin V assay (data not shown). Furthermore, p53-dependent apoptosis was significantly enhanced by TSA treatment in H1299 cells (up to 68% sub-G1); more notably, TSA treatment effectively abrogates PID-mediated downregulation of apoptosis (Fig. 5c), indicating that PID-associated deacetylase activity is required for PID-mediated functions. Together, these data clearly show that PID is involved in the regulation of p53-mediated growth arrest and apoptosis.

Deacetylase activities are required for normal cell-cycle

progression^{15,24}. In addition to causing G1 and G2 phase cell-cycle arrest, deacetylase inhibitors, including sodium butyrate, Trichostatin A and trapoxin, can induce cell transformation reversion, senescence and apoptosis^{15,16,24}. Our data suggest that the biological functions of these deacetylase inhibitors may act on the functional interaction between p53 and PID-associated NuRD complex, at least in part, through inhibition of p53 deacetylation. Acetylation of p53 has been implicated in transcriptional activation⁸⁻¹⁰, senescence²⁵ and chromosome fragility²⁶. The precise mechanism by which PID inhibits p53-mediated functions requires further investigation; however, our results suggest that PID may repress the biological functions of p53 by abrogating its sequence-specific transactivation. Of note, in contrast to p53-mediated growth arrest, both transactivation-dependent and -independent pathways are involved in p53-mediated apoptosis although p53 mutants lacking sequence-specific transactivation function only induce a much slower, less efficient apoptotic response in a cell-type specific manner^{23,27-29}. It is possible that TSA treatment is able to enhance p53 transactivation-dependent apoptosis (this study), but it may have different effects on p53 transactivation-independent (including transrepression-dependent) apoptosis in other cell types³⁰. Expression of MTA1 correlates with the metastatic potential of several human cancer cell lines and tissues¹⁹⁻²¹. The expression level of PID/MTA2 is especially high in rapidly dividing cells⁵. It is very likely that the functional interaction between p53 and PID/MTA2, as well as p53 deacetylation, may play an important role in tumour progression. □

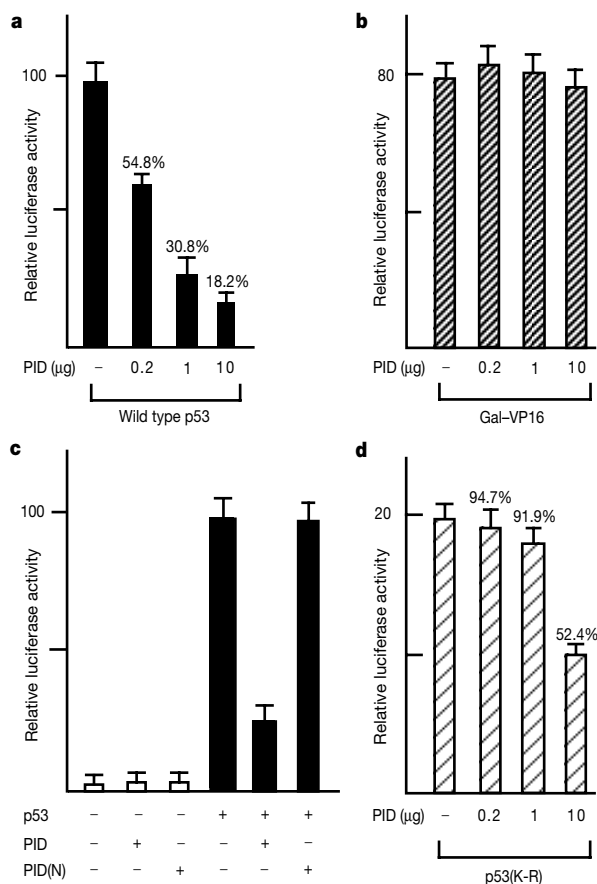


Figure 4 PID repression of transcription activation mediated by wild-type p53, Gal-VP16, or p53(K-R) mutant in cells. **a**, **c**, Wild-type p53; **b**, Gal-VP16; **d**, p53(K-R) mutant. MEF(p53^{-/-}) cells were transiently transfected with p21 mini-Luc reporter construct, CMV-p53wt or CMV-p53(K-R) mutant, and indicated amount of CMV-PID by calcium phosphate precipitation essentially as described¹¹. All transfections were done in duplicate and representative experiments depict the average of three experiments with standard deviations indicated.

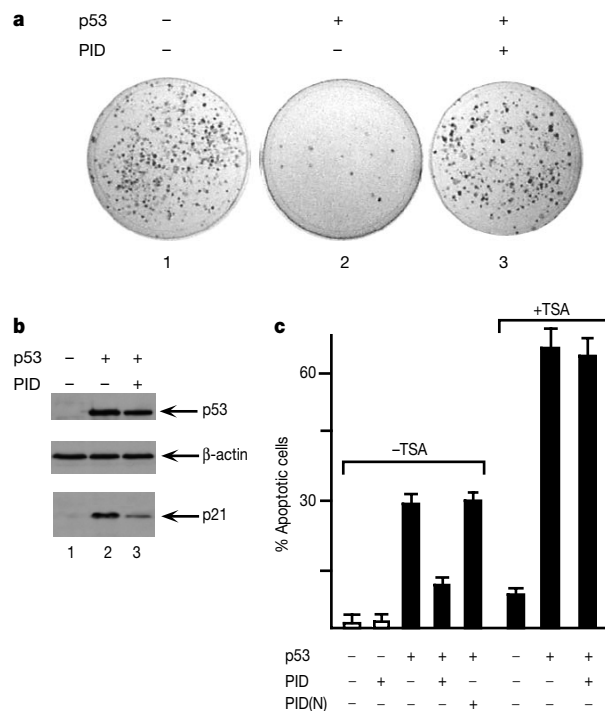


Figure 5 The effects of PID on p53-mediated cell growth arrest, endogenous p21 activation and apoptosis. **a**, SaoS-2 cells were transfected with pTK-hyg, co-transfected with CMV-p53 and pTK-hyg, or co-transfected with CMV-p53, CMV-PID and pTK-hyg. At 24 h post-transfection, cells were split and selected with hygromycin B, and surviving colonies were counted after 2–4 weeks. **b**, Western blot analysis of indicated transfected cell extracts with anti-p53 (DO-1), anti-p21 (Ab1) and anti-β-actin. **c**, H1299 cells were transfected with p53 alone, co-transfected with p53 and PID, or co-transfected with p53 and PID(N). After transfection, cells were fixed, stained for p53 by FITC-conjugated α-p53 antibody, and analysed for apoptotic cells (sub-G1) according to DNA content (PI staining). In the case of TSA treatment, TSA was added to the cells 24 h before collection.

Methods

Preparation of proteins and antibodies

Plasmids encoding PID were created by amplification of the appropriate DNA fragments, including the Flag sequence, followed by subcloning into either pET-11d or pCB6 vectors. The Flag–PID protein was expressed in BL21 (Lys) cells and purified on an M2 agarose affinity column (Sigma). Anti-PID antisera was raised against a peptide at the N terminus (residues 56–71) crosslinked with keyhole limpet hemocyanin (KLH). The rabbit polyclonal antibody specific for p300-mediated acetylated p53 [α -p53CTD(Ac)] was raised against the acetylated human p53 C-terminal peptide (H-S55GQSTSRH55LMF-OH (where 5 represents acetylated lysine)) conjugated through the added C-terminal cysteine to KLH. Antisera from immunized rabbits was first passed through a column coupled with the unacetylated peptide (H-SKKGQSTSRHKKLMF-OH) to deplete antibodies that react with unacetylated p53. This step was repeated several times to remove completely the antibodies against the unacetylated p53. The flow-through anti-sera were further affinity purified by use of acetylated p53 peptide coupled with agarose beads.

Purification of PID from nuclear extracts

The tagged cells were grown in DMEM medium, and nuclear extracts were prepared as described^{17,18}. The nuclear extract prepared from the Flag–HDAC1 cell line was adjusted to 300 mM KCl and 0.2% NP-40 by addition of 3 M KCl and 10% NP-40, and incubated with M2-agarose beads (sigma) at 4 °C for 6 h by rotation. After washing with a buffer BC300 containing 300 mM KCl and 0.2% NP-40, proteins were eluted from beads by incubating at 4 °C for 60 min with BC100/0.2% NP-40 plus 0.5 mg ml⁻¹ Flag peptide. For the differential affinity-binding assay, the Flag–HDAC1 complexes were first dissociated at 2 M urea at 4 °C for 2 h, and then dialysed in a buffer BC200 containing 200 mM KCl and 0.2% NP-40 for affinity binding. The GST–p53 affinity column was prepared by immobilizing the GST–p53 protein on glutathione-sepharose beads under very high stringency conditions (1 M KCl, 1% NP-40 and 0.2% Sarkosyl) to avoid possible leaking of the GST–p53 proteins from the column during elution. The affinity column was first washed with the BC200 buffer, then loaded with dissociated Flag–HDAC1 complex proteins. After extensive washing of the beads with BC300/0.2% NP-40, bound proteins were eluted with a buffer containing 1 M KCl, 1% NP-40 and 0.2% Sarkosyl.

p53 deacetylation assay

Acetylated GST–p53 was prepared by p53 acetylation assay as described¹⁰, and further purified on glutathione-sepharose. The ¹⁴C-labelled acetylated p53 was incubated with purified PID-associated complex or the control eluate at 30 °C for 1 h in a buffer containing 10 mM phosphate (pH 7.0), 100 mM NaCl, 5 mM MgCl₂, 1 mM EDTA, 0.5 mM dithiothreitol, 0.1 mM PMSF and 5% glycerol. The reactions were resolved on SDS–polyacrylamide gel electrophoresis (PAGE) and analysed by Coomassie blue staining and autoradiography.

Received 17 May; accepted 5 October 2000.

- Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
- Prives, C. & Hall, P. A. The p53 pathway. *J. Pathol.* **187**, 112–126 (1999).
- Giacchia, A. J. & Kastan, M. B. The complexity of p53 modulation: emerging patterns from divergent signals. *Genes Dev.* **12**, 2973–2983 (1998).
- Yu, J. *et al.* Identification and classification of p53-regulated genes. *Proc. Natl Acad. Sci. USA* **96**, 14517–14522 (1999).
- Zhang, Y. *et al.* Analysis of NuRD subunits reveals a histone deacetylase core complex and connection with DNA methylation. *Genes Dev.* **13**, 1924–1935 (1999).
- Wade, P. A. *et al.* Mi-2 complex couples DNA methylation to chromatin remodeling and histone deacetylation. *Nature Genet.* **23**, 62–66 (1999).
- Futamura, M. *et al.* Molecular cloning, mapping, and characterization of a novel human gene MTA-1, showing homology to a metastasis associated gene, MTA1. *J. Hum. Genet.* **44**, 52–56 (1999).
- Sakaguchi, K. *et al.* DNA damage activates p53 through phosphorylation-acetylation cascade. *Genes Dev.* **12**, 2831–2841 (1998).
- Liu, L. *et al.* p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. *Mol. Cell Biol.* **19**, 1202–1209 (1999).
- Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90**, 595–606 (1997).
- Gu, W., Shi, X. L. & Roeder, R. G. Synergistic activation of transcription by CBP and p53. *Nature* **387**, 819–823 (1997).
- Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S. & Reinberg, D. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* **95**, 279–289 (1998).
- Xue, Y. *et al.* NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol. Cell* **2**, 851–861 (1998).
- Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E. & Schreiber, S. L. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* **395**, 917–921 (1998).
- Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator RPD3p. *Science* **272**, 408–411 (1996).
- Hassig, C. A. *et al.* A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. *Proc. Natl Acad. Sci. USA* **95**, 3519–3524 (1998).
- Gu, W. *et al.* A novel human SRB/MED-containing cofactor complex (SMCC) involved in transcription regulation. *Mol. Cell* **3**, 97–108 (1999).
- Ito, M. *et al.* Identity between TRAP and SMCC complexes indicates novel pathways for the function of nuclear receptors and diverse mammalian activators. *Mol. Cell* **3**, 361–370 (1999).
- Toh, Y., Pencil, S. D. & Nicolson, G. L. A novel candidate metastasis-associated gene, mta1, differentially expressed in highly metastatic mammary adenocarcinoma cell lines. cDNA cloning, expression, and protein analyses. *J. Biol. Chem.* **269**, 22958–22963 (1994).

- Toh, Y. *et al.* Overexpression of MTA1 gene in gastrointestinal carcinomas: correlation with invasion and metastasis. *Int. J. Cancer* **74**, 459–463 (1997).
- Toh, Y., Kuwano, H., Mori, M., Nicolson, G. L. & Sugimachi, K. Overexpression of metastasis-associated MTA1 mRNA in invasive oesophageal carcinomas. *Br. J. Cancer* **79**, 1723–1726 (1999).
- Lin, J., Chen, J., Elenbaas, X. & Levine, A. J. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1b 55-kD protein. *Genes Dev.* **8**, 1235–1246 (1994).
- Jimenez, G. S. *et al.* A transactivation-deficient mouse model provides insights into Trp53 regulation and function. *Nature Genet.* **26**, 37–43 (2000).
- Yoshida, M., Horinouchi, S. & Beppu, T. Trichostatin A and trapoxin: Novel chemical probes for the role of histone acetylation in chromatin structure and function. *BioEssays* **5**, 423–430 (1995).
- Pearson, M. *et al.* pML regulates p53 acetylation and premature senescence induced by oncogenic Ras. *Nature* **406**, 207–210 (2000).
- Yu, A., Fan, H., Lao, D., Bailey, A. D. & Weiner, A. M. Activation of p53 or loss of the Cockayne syndrome group B repair protein causes metaphase fragility of human U1, U2, and 5S genes. *Mol. Cell* **5**, 801–810 (2000).
- Amundson, S. A., Nysers, T. G. & Flornace, A. J. Roles for p53 in growth arrest and apoptosis: putting on the brakes after genotoxic stress. *Oncogene* **17**, 3287–3299 (1998).
- Attardi, L. D., Lowe, S. W., Brugarolas, J. & Jacks, T. Transcriptional activation by p53, but not induction of the p21 gene, is essential for oncogene-mediated apoptosis. *EMBO J.* **15**, 3693–3701 (1996).
- Haupt, Y., Rowan, S., Shaulian, E., Vowsden, K. H. & Oren, M. Induction of apoptosis in HeLa cells by trans-activation deficient p53. *Genes Dev.* **9**, 2170–2183 (1995).
- Murphy, M. *et al.* Transcriptional repression by wild-type p53 utilizes histone deacetylase, mediated by interaction with mSin3a. *Genes Dev.* **13**, 2490–2501 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Dalla-Favera, R. Baer and B. Tycko for critical discussions, and use of laboratory space and reagents; K. Vowsden, B. Vogelstein, A. Levine, M. Oren, Y. Xiong, C. Hassig, S. L. Schreiber, S. Chellappan, G. Lozano and P. P. Pandolfi for antibodies, cell lines and plasmids; J. Qin, W. Wang and Y. Zhang for help; G. Cattoretti and H. Niu for suggestions in apoptosis assays; the sequencing facility of Columbia University Cancer Center for sequencing; F. Huang, M. Li, A. Nikolaev and N. A. Papanikolaou for sharing unpublished data and critical comment; and R.G. Roeder for continuous support and encouragement. This work was supported in part by grants from the NIH/NCI, the American Cancer Society and the Herbert Irving Comprehensive Cancer Center to W.G.

Correspondence and requests for material should be addressed to W.G. (e-mail: wg8@columbia.edu). The GenBank accession number for the PID sequence is AF295807.

Insights into SCF ubiquitin ligases from the structure of the Skp1–Skp2 complex

Brenda A. Schulman*, Andrea C. Carrano†, Philip D. Jeffrey*, Zachary Bowen*, Elspeth R. E. Kinnucan*, Michael S. Finnin*§, Stephen J. Elledge‡§, J. Wade Harper‡, Michele Pagano† & Nikola P. Pavletich*§

* Cellular Biochemistry and Biophysics Program, and § Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

† Department of Pathology and Kaplan Comprehensive Cancer Center, NYU Medical Center, New York, New York 10016, USA

‡ Verna and Marrs McLean Department of Biochemistry and § Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA

F-box proteins are members of a large family that regulates the cell cycle, the immune response, signalling cascades and developmental programmes by targeting proteins, such as cyclins, cyclin-dependent kinase inhibitors, I κ B α and β -catenin, for ubiquitination (reviewed in refs 1–3). F-box proteins are the substrate-recognition components of SCF (Skp1–Cullin–F-box protein) ubiquitin-protein ligases^{4,5}. They bind the SCF constant catalytic core by means of the F-box motif interacting with Skp1, and they bind substrates through their variable protein–protein

interaction domains⁶. The large number of F-box proteins is thought to allow ubiquitination of numerous, diverse substrates⁶. Most organisms have several Skp1 family members, but the function of these Skp1 homologues and the rules of recognition between different F-box and Skp1 proteins remain unknown. Here we describe the crystal structure of the human F-box protein Skp2 bound to Skp1. Skp1 recruits the F-box protein through a bipartite interface involving both the F-box and the substrate-recognition domain. The structure raises the possibility that different Skp1 family members evolved to function with different subsets of F-box proteins, and suggests that the F-box protein may not only recruit substrate, but may also position it optimally for the ubiquitination reaction.

The ~40-residue F-box motif is the common feature of F-box proteins⁶. Variable 'linker' sequences connect F-boxes to protein-protein interaction modules, such as leucine-rich repeats (LRRs) or WD-40 repeats, which are thought to bind directly to substrates. The human F-box protein Skp2 has LRRs, and it regulates the G1/S transition in mammalian cells^{7,8} by controlling the abundance of the cyclin-dependent kinase inhibitor, p27^{Kip1} (refs 8–11). Here we describe the structure of a Skp1–Skp2 complex that lacks an

amino-terminal 100-residue region of unknown function in Skp2, but that promotes ubiquitination of p27^{Kip1} *in vitro* to an extent comparable to that of full-length Skp2 (ref. 9).

The structure of the Skp1–Skp2 complex resembles a sickle, with Skp1 and the Skp2 F-box representing the handle and the LRR domain representing the curved blade (Fig. 1). Skp1 has a ~125-residue N-terminal domain with an α/β structure similar to that of the BTB/POZ domain fold¹², but with a helical insertion (H4) and a two-helix carboxy-terminal extension (H7 and H8). The last two helices of the BTB/POZ fold (H5 and H6) and the two helices of the C-terminal extension form the Skp2-binding site (Figs 1 and 2). The structure of the Skp2 F-box consists of three helices, with the H1 helix packing orthogonally with the H2–H3 antiparallel helix pair. The ~70-residue 'linker' following the F-box adopts the structure of three non-canonical LRRs that are contiguous with the seven LRRs predicted from the amino-acid sequence⁷. The ten LRRs, each consisting of a β -strand and an α -helix, pack into a curved structure¹³. They form a single structural domain that is attached directly to the F-box. After the tenth LRR, the ~30-residue C-terminal tail of Skp2 extends back toward the first LRR, packs loosely in the concave surface of the LRR domain and terminates in a short β -strand inserted at the interface between Skp1 and Skp2 (Fig. 1).

The extensive Skp1–Skp2 interface is predominantly hydrophobic and buries a total of 2,980 Å² (Figs 1, 3). The interface is interdigitated, with alternating structural elements from Skp1 and Skp2 packing into four layers. The F-box and the Skp1 H8 helix, which are the middle two layers, are sandwiched between the H5, H6 and H7 helices of Skp1 on one side and the first LRR and C-terminal strand of Skp2 on the other side. The interface is best described in two parts from the perspective of the F-box. In one part, one face of the F-box packs with the H5, H6 and H7 helices of Skp1. This is the core interface, as it accounts for two-thirds of the surface area buried (2,050 Å²) and contains residues conserved in all Skp1 and F-box protein family members (Fig. 3). In the second part, the opposite face of the F-box packs with the H8 helix of Skp1 and also with the first LRR and C-terminal strand of Skp2, burying 930 Å² of the surface area. We will refer to this as the variable part of the interface, because the structural elements that the F-box binds here are conserved in Skp1 and Skp2 orthologues, but are not conserved in other F-box and Skp1 family members (such as the 17 Skp1 and 215 F-box sequences in the *Caenorhabditis elegans* genome; Fig. 3). In particular, the Skp1 sequence corresponding to the H8 helix, although highly conserved in Skp1 orthologues (Fig. 2; see also Supplementary Information), is variable in length and sequence among other Skp1 family members in worms, flies and plants. Furthermore, the C-terminal tails of other LRR-containing F-box proteins are also variable in length and sequence, and the portion of the interface corresponding to the first LRR would be different in

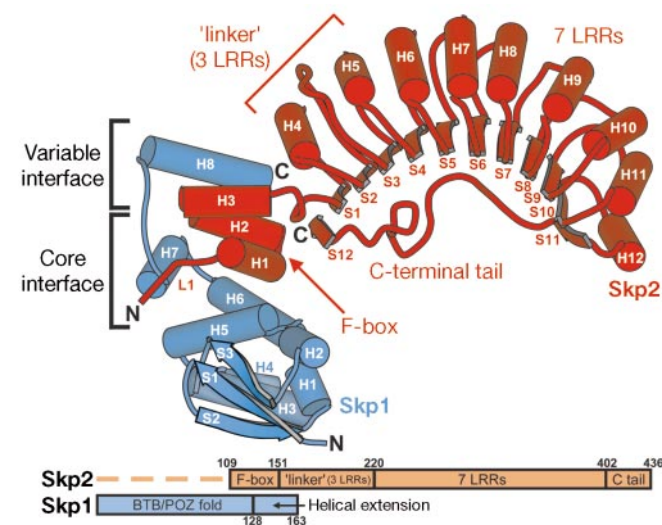


Figure 1 Structure of the Skp1–Skp2 complex. Skp1 is shown in blue and Skp2 is shown in red. The boundaries of the BTB/POZ fold, the C-terminal helical extension of Skp1 and of the F-box, the three non-canonical LRRs, the seven canonical LRRs and the C-terminal tail of Skp2 are shown in the diagram below the structure. The 100-residue N-terminal Skp2 region missing from the crystallized protein is indicated (dashed line). The second LRR has a partially disordered loop instead of the helix characteristic of LRRs.

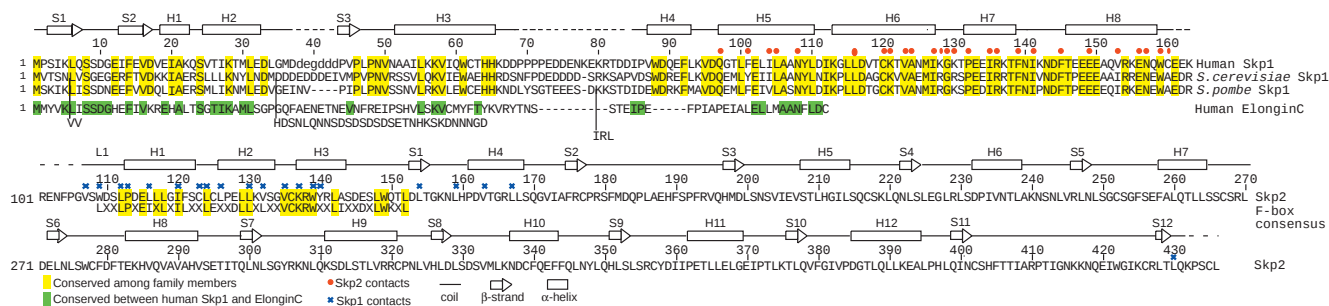


Figure 2 Structural elements and conservation in Skp1 and Skp2. Top, sequence identity between human Skp1 and its orthologues from *S. cerevisiae* and *Schizosaccharomyces pombe* (yellow), and its distant relative, human ElonginC (green). ElonginC is shorter than Skp1, lacking the H4 helix, the last helix of the BTB/POZ fold (H6) and the two-helix

C-terminal extension (H7 and H8) of Skp1. Red dots indicate Skp1 residues that contact Skp2; dashed lines indicate disordered segments; lower-case letters denote the internal deletion. Bottom, secondary structure of Skp2, with identity between the Skp2 F-box and the F-box consensus sequence^{1,6} (yellow). Blue stars indicate residues that contact Skp1.

non-LRR-containing F-box proteins.

In the core part of the interface, the H5, H6 and H7 helices of Skp1 and the H1, H2 and H3 helices of the F-box pack together in a right-handed superhelical arrangement. The most extensive contacts are made by the L1 loop and H1 helix of the F-box. From the F-box L1 loop, Trp 109, Leu 112 and Pro 113 contact Leu 100, Phe 101, Ile 104, Leu 105, Val 123 and Phe 139 of Skp1 (Figs 2 and 3). From the F-box H1 helix, Leu 116, Ile 120 and Leu 124 contact Ile 104, Leu 105, Leu 116, Cys 120, Val 123 and Ile 127 of Skp1 (Figs 2 and 3). The H2 and H3 F-box helices contribute fewer contacts (Val 132, Val 135 and Trp 139; Figs 2 and 3). Several of the Skp1-contacting F-box residues also have structural functions: Pro 113, a hallmark of the F-box motif, helps to initiate the H1 helix; and Leu 124 and Trp 139 both contribute to the packing of the F-box helices.

The large buried surface area and the extensive intermolecular contacts at the core interface suggest that it is the primary binding site, and explain how an isolated F-box can bind to Skp1 (refs 4, 6, 14, 15). Indeed, the core interface is essentially the same as in the crystal structure of the isolated F-box bound to a truncated Skp1, which lacks the H8 helix (Skp1 Δ H8), determined in this study (Table 1). Many of the mutations that disrupt function map to residues in the core^{6,16,17} (corresponding to Gly 129, Arg 136 and Ile 141 of Skp1, and to Leu 112, Pro 113, Ile 120, Phe 121 and Leu 124 of the Skp2 F-box). The fact that the F-box and Skp1 residues here are highly conserved in their respective families (Fig. 3) further suggests that the core interface structure will be similar in all combinations of F-box protein and Skp1 family members.

In the variable part of the interface, Skp2 has a large groove formed by residues from the H2 and H3 helices of the F-box

(Leu 126, Leu 129, Leu 130, Lys 137, Tyr 140 and Trp 149), the first LRR (Leu 152, Leu 154, Leu 159, Val 163 and Leu 167) and the C-terminal strand (Leu 428 and Leu 430). This groove binds the H8 helix of Skp1 through a combination of van der Waals and hydrogen-bond contacts (to Phe 145, Glu 149, Val 153, Glu 156, Asn 157, Trp 159 and Cys 160 of Skp1; Fig. 3). The most important Skp2 residue in this portion of the interface is Trp 149, which is the only conserved F-box residue here (Fig. 2). Although the side chain of Trp 149 does not contact Skp1, it has a central organizing function at the bottom of the Skp2 groove, packing with residues from the F-box, the first LRR and the C-terminal strand of Skp2. The most important Skp1 residue in this portion of the interface is Trp 159, which is near the end of the Skp1 H8 helix. It binds in the deepest portion of the Skp2 groove, contacting the first LRR, the C-terminal strand and the F-box, and serves as the linchpin for this part of the interface. The importance of the variable interface is highlighted by our finding that the Skp1 H8 helix and Trp 159 are needed for Skp1 function *in vivo*. When we made a mutation corresponding to Trp 159 to alanine or deleted the H8 helix of the *Saccharomyces cerevisiae* Skp1 orthologue, these mutants failed to complement the temperature sensitivity of the *skp1-11* strain (see Supplementary Information).

To test the contribution that this unexpected variable interface makes to the stability of the complex, we used an *in vitro* competition assay to compare the dissociation rates of the Skp2–Skp1 complex with those of the F-box–Skp1 Δ H8 complex, which has a core interface structure identical to that of the Skp1–Skp2 complex but that lacks all of the elements unique to the variable interface. The Skp1–Skp2 complex has a half-life greater than 9 h, but the F-box–Skp1 Δ H8 complex has a half-life of less than 30 min, indicating that the variable part of the interface contributes substan-

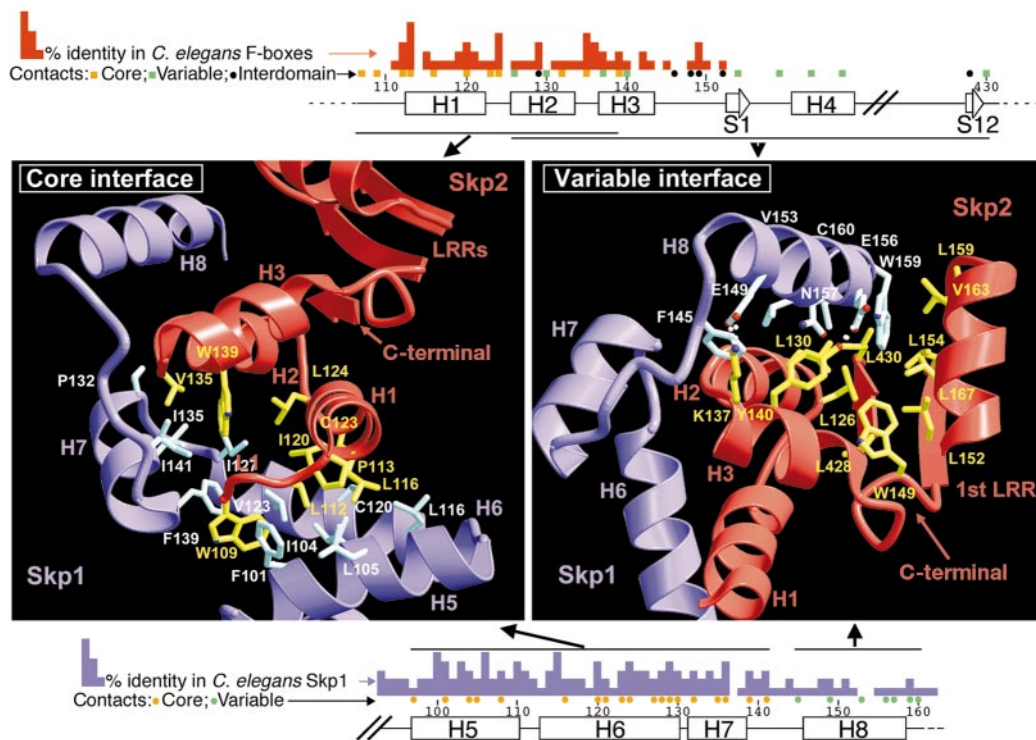


Figure 3 Bipartite Skp1–Skp2 interface. Left, the core interface involves H5, H6 and H7 from Skp1, and L1, H1, H2 and H3 from the F-box. Secondary structures of Skp1 (purple) and Skp2 (red), and side chains of Skp1 (cyan) and Skp2 (yellow) are shown. Right, the variable interface involves H8 from Skp1 and the F-box, first LRR and C-terminal strand from Skp2. White dashes represent hydrogen bonds. Top, conservation of Skp2 residues that make intermolecular contacts to Skp1 in either the core (orange squares) or variable (green squares) parts of the interface, or that make interdomain contacts between the

F-box and substrate-binding domain (black dots). Red histogram represents identity among 20 *C. elegans* F-box sequences. Bottom, conservation of Skp1 residues that contact Skp2 in either the core (orange dots) or variable (green dots) parts of the interface (bottom). Violet histogram represents identity among the 17 *C. elegans* Skp1 homologues. Similar histograms are obtained for F-box and Skp1 family members from *D. melanogaster*.

tially to the stability of the complex. This is consistent with studies suggesting that regions outside the F-box enhance association with Skp1 (refs 14, 15, 18).

Although all but one of the Skp1 and Skp2 residues at the second part of the interface are variable in their respective families, several observations suggest that an analogous interface (involving the substrate-binding domain, the F-box and a C-terminal portion of Skp1) will form among most combinations of F-box protein and Skp1 family members, even if the precise structures and residues differ. First, Trp 149 of Skp2 is conserved in both LRR and in non-LRR F-box proteins⁶, suggesting that it will couple the F-box and substrate-binding domain in general. Second, the other F-box residues that pack with the Skp1 H8 helix, although not conserved, maintain their overall hydrophobic/hydrogen bonding nature in other F-box proteins. Third, Trp 159 of Skp1 is invariant in orthologues from yeast to humans, suggesting that it will also contact the F-box and substrate-binding domains of non-LRR F-box proteins. Other Skp1 family members contain hydrophobic amino acids near their C termini that may serve analogous functions.

The crystal structure provides a starting point for understanding specificity between different Skp1 and F-box family members. Although *S. cerevisiae* contains a single Skp1, the genomes of higher eukaryotes encode several proteins with homology to most of the Skp1 sequence (S1 to H7). The genomes of *C. elegans* and *Drosophila melanogaster* contain at least 17 and 7 Skp1-like sequences, respectively^{19,20}. Although it is not yet known whether these proteins function in SCF-like ubiquitin-protein ligases, our structure suggests that they are close structural homologues of Skp1, except for the C-terminal H8 helix. For example, the cluster of Trp 61, His 65, Trp 88 and Asp 89, which stabilizes the structure through aromatic stacking and a buried hydrogen bond, is invariant in all the fly sequences and highly conserved in the worm sequences, but is entirely absent from other BTB/POZ fold proteins.

Our structure suggests that these Skp1 family members are likely to be functional homologues as well. The strong conservation of the residues in the core portion of the Skp1–Skp2 interface indicates that most of these Skp1 family members should be able to bind to a consensus F-box (Fig. 3). In theory, promiscuous binding between F-box and Skp1 family members would result in combinatorial

diversity in the selection and regulation of SCF targets³. However, our data suggest that, to provide a complementary structure at the variable portion of the interface, Skp1 family members with divergent C-terminal sequences will preferentially associate with specific subsets of F-box proteins, generating specificity in targeting proteins for degradation.

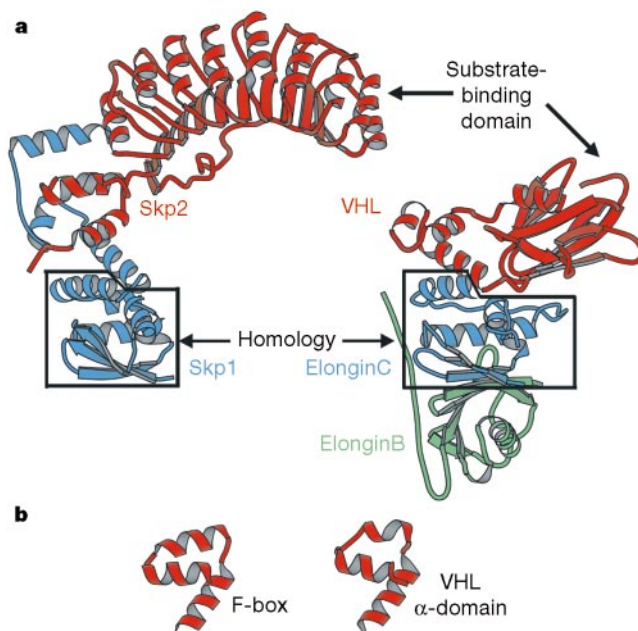


Figure 4 Comparison of the Skp2–Skp1 and VHL–ElonginC–ElonginB complexes. **a**, Homologous portions of Skp1 and ElonginC are aligned and boxed. Skp2 and VHL are red, Skp1 and ElonginC are blue and ElonginB is green. The LRRs of Skp2 (refs 4, 8, 26, 27) and the β-domain of VHL are thought to bind substrate²³. Owing to its unique C terminus, Skp1 binds the F-box differently from the way that ElonginC binds VHL. The arrangement of helices in the two interfaces is similar, although the helices come from non-corresponding members of the complexes (H1 and H2 from VHL superimpose with H6 and H7 from Skp1, not helices from Skp2). **b**, The ElonginC-binding region of VHL resembles the three-helix cluster structure of the F-box.

Table 1 Statistics from the crystallographic analysis

Protein complex	Skp1–Skp2*				Skp1ΔH8–F-box + LRRs†			Skp1ΔH8–F-box‡		
	Native	Native	Au	Hg	Native	Pt	Hg	Native	Pt	Hg
Beamline	CHESSA1	NLSX9B	NLSX9B	NLSX9B	NLSX9B	–	–	NLSX9B	–	–
Space group	C2	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁	P2 ₁
Resolution (Å)	2.8	2.9	3.2	3.5	1.8	2.7	2.4	1.8	2.7	2.4
Observations	598,299	178,486	122,143	87,911	206,639	48,153	65,198	206,639	48,153	65,198
Unique reflections	108,372	23,509	19,152	15,568	29,068	9,847	13,981	29,068	9,847	13,981
Completeness (%) (last shell)	93.7 (93.8)	96.9 (98.7)	95.2 (98.1)	92.6 (95.6)	84.4 (47.1)	96.2 (84.5)	93.5 (83.9)	84.4 (47.1)	96.2 (84.5)	93.5 (83.9)
R _{sym} (%) (last shell)	7.4 (19.4)	8.5 (38.8)	7.6 (36.3)	15.8 (49.7)	3.7 (12.4)	6.0 (18.7)	7.1 (21.8)	3.7 (12.4)	6.0 (18.7)	7.1 (21.8)
I/σ (last shell)	17.9 (5.1)	13.9 (1.9)	18.2 (3.5)	7.9 (2.1)	31.9 (6.7)	19.7 (4.0)	27.6 (6.1)	31.9 (6.7)	19.7 (4.0)	27.6 (6.1)
MIR analysis										
Phasing power			0.72	0.97		1.12	1.45			
R _{culis}			0.90	0.87		0.83	0.74			
Refinement statistics										r.m.s.d.
Protein complex	Resolution (Å)	Reflections	Total atoms	Water atoms	R-factor (%) (last shell)	R _{free} (%) (last shell)	Bonds (Å)	Angles (°)	B factor (Å)	
Skp1–Skp2*	8–2.8	89,527 (F > 2σ)	29,168	–	27.5 (35.3)	31.4§ (42.9)	0.010	1.56	2.72	
Skp1ΔH8–F-box + LRRs†	20–2.9	23,508 (F > 0)	6,035	–	24.6 (37.8)	28.7 (38.9)	0.012	1.98	2.33	
Skp1ΔH8–F-box‡	8–1.8	27,706 (F > 0)	3,336	842	21.8 (28.6)	27.4 (34.3)	0.015	1.89	4.02	

R_{sym} = $\sum_i \sum_j |I_{ij} - \bar{I}_i| / \sum_i \sum_j I_{ij}$ for the intensity (I) of i observations of reflection h . Phasing power = $\langle F_o \rangle / E$, where $\langle F_o \rangle$ is the r.m.s. heavy atom structure factor and E is the residual lack of closure error. R_{culis} is the mean residual lack of closure error divided by the dispersive difference. R factor = $\sum_i |F_{obs} - F_{calc}| / \sum_i F_{obs}$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} = R factor calculated using 5% of the reflection data chosen randomly and omitted from the start of refinement. r.m.s.d., root-mean-square deviations from ideal geometry and variation in the B factor of bonded atoms.

* Skp1(1–37;44–68;84–163)–Skp2(101–436).

† Skp1(1–37;44–147)–Skp2(101–153;194–436).

‡ Skp1(1–37;44–147)–Skp2(101–153).

§ Because of non-merohedral twinning, a minor portion of the data contains contributions from a secondary lattice.

In addition to SCF complexes and their close family members, the VHL von Hippel–Lindau tumor suppressor–ElonginC–ElonginB complex may function in an ubiquitin–protein ligase analogous to the SCF^{21,22}. Comparison of the Skp2–Skp1 complex with the structure of the VHL–ElonginC/ElonginB complex²³, however, shows that Skp1 and ElonginC bind their respective Skp2 and VHL partners in different ways (Fig. 4a). Although the structure of the 112-residue ElonginC is similar to the N-terminal two-thirds of Skp1 (C α r.m.s deviation of 1.27 Å for 74 out of 112 residues), ElonginC lacks the H6, H7 and H8 helices of Skp1 that form the bulk of the Skp2-binding site (Fig. 4a).

The Skp2–Skp1 structure also reveals similarities with the VHL–ElonginC–ElonginB system, which indicate an evolutionary relationship. It reveals (1) that the α -domain of VHL, which mediates most of the binding to ElonginC, has a structure similar to that of the F-box of Skp2 (Fig. 4b); (2) that as in the Skp2 structure, the VHL β -domain, which was proposed to be a substrate-binding domain²³, is tightly coupled to the α -domain; and (3) that the relative positions of the possible substrate-binding domains of Skp2 and VHL and the homologous portions of Skp1 and ElonginC, which may be involved in binding the cullin subunit, are similar (Fig. 4a).

It is not yet understood how the SCF ubiquitinates its substrates. It is possible that the main function of the F-box protein is to recruit the substrate to raise its effective concentration. However, the rigid coupling between the Skp2 substrate-binding domain and Skp1 indicates that the F-box protein may also contribute to the optimal positioning of the substrate for ubiquitination. Although there are no known consensus sequences for ubiquitin-accepting lysines, p27^{Kip1} and I κ B α are both ubiquitinated at specific lysine side chains close to their F-box protein-binding sites^{24,25}. The selection of lysines for ubiquitination may be determined by the spatial and stereochemical constraints imposed by the tight coupling of the substrate-binding domain with Skp1 and its associated SCF catalytic core.

We do not know how the Skp2 LRR domain binds substrates, but in the LRR-containing proteins RNase inhibitor and U2A', the concave face of the LRR β -sheet is involved in protein–protein interactions^{26,27}. In Skp2, most of the concave face of the LRR domain is filled with a loosely packed 25-residue portion of the Skp2 tail, raising the possibilities that the tail may participate in or regulate substrate binding, or that it may itself be a substrate³. On this last possibility, we note that Skp2 protein levels oscillate during the cell cycle^{7,18}. Notably, three lysine residues, Lys 416, Lys 417 and Lys 425, in this part of the Skp2 tail protrude away from the structure and may be sites for ubiquitination.

The structure of the Skp1–Skp2 complex provides the first structural insights into the largest class of ubiquitin–protein ligases. It reveals both a general mechanism for F-box protein recruitment to SCF ligases and a second binding site that provides a basis for specificity among family members, and it suggests that the F-box protein may also help position the substrate for the ubiquitination reaction. □

Methods

Protein expression and purification

Three complexes are part of this study. Skp1 Δ H8–F-box contains the Skp2 F-box (residues 101–153) and a truncated Skp1 (1–147). Skp1 Δ H8–F-box + LRRs contains the same Skp1 mutant and the F-box attached to the seven predicted LRRs (residues 194–409) through an engineered linker. The F-box packs with the third LRR in this structure, but in an artificial manner. We used this complex to determine the structure of the LRR domain to facilitate the structure determination of the Skp1–Skp2 complex that is described here (Table 1). All three structures lack a six-residue internal loop of Skp1 (residues 38–43), which is in a region that we determined to be unstructured on the basis of proteolytic digestion, and is absent from several Skp1 orthologues. The Skp1–Skp2 complex also lacks a protease-sensitive Skp1 loop (residues 71–82) that was present in the Skp1 Δ H8–F-box and Skp1 Δ H8–F-box + LRRs complexes but that was disordered in these crystal structures. Deletion of either Skp1 loop does not affect binding to either Skp2 or Cull1, as assayed by glutathione S-transferase pull-down experiments.

Human Skp2 and Skp1, or their fragments, were cloned as a dicistronic message into the

pGEX 4T1 plasmid (Pharmacia)²³, with the GST–Skp2 fusion coding region preceding that of Skp1. Following co-expression in BL21(DE3) at 25 °C, the complexes were purified by glutathione affinity, anion exchange and gel-filtration chromatography and were concentrated by ultrafiltration to ~30–90 mg ml⁻¹ in 50 mM Tris–HCl, 150–200 mM NaCl and 5 mM dithiothreitol (DTT), pH 7.6.

Crystallization and structure determination

We grew crystals using the hanging-drop vapour-diffusion method. Crystals of Skp1 Δ H8–F-box grew in 36–38% PEG 1500, 0.1 M sodium acetate, pH 4.6, 0.2 M ammonium acetate and 5 mM DTT at 25 °C. They form in the space group P2₁, with $a = 46.5$, $b = 41.6$, $c = 87.2$ Å and $\beta = 93.4^\circ$, and have two complexes in the asymmetric unit. Heavy atom soaks were performed in 38% PEG 1500, pH 6.2, supplemented with 3 mM K₂PtCl₆ (3.5 h) or 2.75 mM thimerosal (4 h). Crystals were flash frozen in crystallization buffer containing 43% PEG 1500.

Crystals of Skp1 Δ H8–F-box + LRRs grew in 16–18% PEG 4000, 0.1 M Tris–HCl, 0.2 M NaCl and 5 mM DTT, pH 7.4, at 4 °C, by microseeding. They form in the space group P2₁2₁2₁ with $a = 72.6$, $b = 87.0$ and $c = 178.3$ Å, and have two complexes in the asymmetric unit. Heavy atom soaks were performed in 19% PEG 4000, pH 7.4, supplemented with 1 mM KAu(CN)₂ (2.5 h) or 0.5 mM thimerosal (4 h). Crystals were flash frozen in crystallization buffer supplemented with 30% ethylene glycol.

Crystals of the Skp1–Skp2 complex were grown in 15–16% PEG 4000, 0.1 M Tris–HCl, 0.2 M ammonium acetate, 10 mM DTT, pH 8.5, at 4 °C, by microseeding and macroseeding. They form in the space group C2, with $a = 262.7$, $b = 148.2$, $c = 133.3$ Å and $\beta = 120.0^\circ$, and have eight complexes in the asymmetric unit. These crystals had non-merohedral twinning, producing two closely spaced non-identical diffraction lattices. The combined lattices could be integrated together to yield unit-cell parameters and symmetry consistent with space group I222 ($a = 133.4$, $b = 148.2$, $c = 227.6$ Å). The individual lattices are not consistent with I222 unit cell parameters (β becomes 90.5°), however, and when integrated separately in space group P1, the data have one exact two-fold symmetry axis (R_{sym} 4.5%) along the direction of the b axis and two pseudo-two-fold symmetry axes along a and c (R_{sym} 25–30%). With only one crystallographic two-fold axis, the appropriate space group is C2 with a doubling of the I222 asymmetric unit. Most of the spots could be integrated separately with DENZO²⁸, but presumably some reflections contain minor portions of the secondary lattice. The crystals were flash frozen in crystallization buffer containing 19% PEG 4000 and 30% ethylene glycol.

We processed all data with DENZO²⁸ and SCALEPACK²⁸. For the Skp1 Δ H8–F-box and the Skp1 Δ H8–F-box + LRRs complexes, multiple isomorphous replacement (MIR) phases were calculated with MLPHARE²⁹ and were improved by solvent flattening and two-fold non-crystallographic averaging with DM²⁹ at 2.7 and 3.5 Å, respectively. For the Skp1 Δ H8–F-box + LRRs complex, the Skp1–F-box complex was located by molecular replacement with AMORE²⁹ using the Skp1 Δ H8–F-box structure as a search model and the 7-LRR domain was built using the MIR maps. For the structure determination of the Skp1–Skp2 complex, the seven LRRs were first located by molecular replacement using the LRR structure of Skp1 Δ H8–F-box + LRRs as a search model. Phases derived from this model were used as the starting point for iterative, eight-fold multidomain non-crystallographic averaging. This allowed for the location of Skp1 and the F-box, and the building of the regions missing in the other two structures. Representative electron density is shown in the Supplementary Information.

All three structures were refined with CNS³⁰. The Skp1 Δ H8–F-box + LRRs and Skp1–Skp2 structures were refined with two-fold and eight-fold non-crystallographic restraints, respectively, including individual temperature-factor refinement. In the Skp1–Skp2 structure, residue 1, the second engineered linker, residues 161–163 of Skp1 and residues 101–106, 187–189 and 432–436 of Skp2 have no electron density. In addition, residues 184–192 from the second LRR of Skp2 have high temperature factors and are presumably partially disordered. The fraction of residues in the most favoured and additionally allowed regions of the Ramachandran plot, respectively, are 93% and 7% for the Skp1 Δ H8–F-box complex, 84% and 16% for the Skp1 Δ H8–F-box+LRRs, and 82% and 18% for the Skp1–Skp2 complex. None of these structures has any residues in a disallowed region.

Yeast temperature-sensitivity complementation assay

The *skp1-11* strain has been described⁶. *S. cerevisiae* SKP1 and mutants were expressed under control of the actin promoter in pJP190 (a gift from Z. Zaman), which carries the His marker. Experiments were carried out on SC–His plates. Six independent clones from each transformation were re-streaked, with three incubated each at 25 °C and 37 °C. We analysed 2 μ g protein from lysates of yeast grown at 25 °C in SC + glucose–His for the presence of Skp1 (ref. 6).

Skp1 dissociation assay

We mixed 280 μ M of the various GST–Skp2–Skp1 complexes with 7.5 mM His–Skp1 (ref. 9). Glutathione S-transferase pull-down assays were performed after 0, 0.5, 1, 2, 4, 9 and 24 h. Dissociation was monitored by gel electrophoresis, based on exchange of Skp1 in the starting complex with the slower migrating His–Skp1.

Received 6 June; accepted 4 September 2000.

- Patton, E. E., Willems, A. R. & Tyers, M. Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. *Trends Genet.* **14**, 236–243 (1998).
- Koepp, D. M., Harper, J. W. & Elledge, S. J. How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell* **97**, 431–434 (1999).
- Deshais, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annu. Rev. Cell Dev. Biol.* **15**, 435–467 (1999).

4. Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. & Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209–219 (1997).
5. Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 221–230 (1997).
6. Bai, C. *et al.* SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**, 263–274 (1996).
7. Zhang, H., Kobayashi, R., Galaktionov, K. & Beach, D. p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S phase kinase. *Cell* **82**, 915–925 (1995).
8. Sutterlüty, H. *et al.* p45SKP2 promotes p27Kip1 degradation and induces S phase in quiescent cells. *Nature Cell Biol.* **1**, 207–214 (1999).
9. Carrano, A. C., Eytan, E., Hershko, A. & Pagano, M. SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nature Cell Biol.* **1**, 193–199 (1999).
10. Tsvetkov, L. M., Yeh, K. H., Lee, S. J., Sun, H. & Zhang, H. p27(Kip1) ubiquitination and degradation is regulated by the SCF(Skp2) complex through phosphorylated Thr187 in p27. *Curr. Biol.* **9**, 661–664 (1999).
11. Nakayama, K. *et al.* Targeted disruption of Skp2 results in accumulation of cyclin E and p27(Kip1), polyploidy and centrosome overduplication. *EMBO J.* **19**, 2069–2081 (2000).
12. Aravind, L. & Koonin, E. V. Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain. *J. Mol. Biol.* **285**, 1353–1361 (1999).
13. Kobe, B. & Deisenhofer, J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366**, 751–756 (1993).
14. Li, F. N. & Johnston, M. Grr1 of *Saccharomyces cerevisiae* is connected to the ubiquitin proteolysis machinery through Skp1: coupling glucose sensing to gene expression and the cell cycle. *EMBO J.* **16**, 5629–5638 (1997).
15. Zhou, P. & Howley, P. M. Ubiquitination and degradation of the substrate recognition subunits of SCF ubiquitin-protein ligases. *Mol. Cell* **2**, 571–580 (1998).
16. Connelly, C. & Hieter, P. Budding yeast SKP1 encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. *Cell* **86**, 275–285 (1996).
17. Russell, I. D., Grancell, A. S. & Sorger, P. K. The unstable F-box protein p58-Ctf13 forms the structural core of the CBF3 kinetochore complex. *J. Cell Biol.* **145**, 933–950 (1999).
18. Lisztwan, J. *et al.* Association of human CUL-1 and ubiquitin-conjugating enzyme CDC34 with the F-box protein p45(SKP2): evidence for evolutionary conservation in the subunit composition of the CDC34-SCF pathway. *EMBO J.* **17**, 368–383 (1998).
19. Costanzo, M. C. *et al.* The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**, 73–76 (2000).
20. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
21. Lonergan, K. M. *et al.* Regulation of hypoxia-inducible mRNAs by the von Hippel-Lindau tumor suppressor protein requires binding to complexes containing elongins B/C and Cul2. *Mol. Cell. Biol.* **18**, 732–41 (1998).
22. Kamura, T. *et al.* Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science* **284**, 657–661 (1999).
23. Stebbins, C. E., Kaelin, W. G. Jr & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455–461 (1999).
24. Shirane, M. *et al.* Down-regulation of p27(Kip1) by two mechanisms, ubiquitin-mediated degradation and proteolytic processing. *J. Biol. Chem.* **274**, 13886–13893 (1999).
25. Spencer, E., Jiang, J. & Chen, Z. J. Signal-induced ubiquitination of I κ B α by the F-box protein Slimb/ β -TrCP. *Genes Dev.* **13**, 284–294 (1999).
26. Kobe, B. & Deisenhofer, J. A structural basis of the interactions between leucine-rich repeats and protein ligands. *Nature* **374**, 183–186 (1995).
27. Price, S. R., Evans, P. R. & Nagai, K. Crystal structure of the spliceosomal U2B''–U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650 (1998).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **176**, 307–326 (1997).
29. Collaborative Computational Project B. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
30. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary information is available on *Nature's* World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank H. Erdument-Bromage for N-terminal sequence and mass spectroscopic analysis; P. Sorger for communication of results before publication; A. Ansari, P. Murray and members of the Pavletich lab for discussions; C. Murray for administrative assistance; and the staff of the National Synchrotron Light Source X9B beamline and of the Cornell High Energy Synchrotron Source MacChess for help with data collection. B.A.S. was supported by a Foundation for Advanced Cancer Studies fellowship from the Life Sciences Research Foundation. This work was supported by the NIH, the Howard Hughes Medical Institute, the Dewitt Wallace Foundation, the Samuel and May Rudin Foundation, the Human Frontiers Science Program and the Welch Foundation.

Correspondence and requests for materials should be addressed to N.P.P. (e-mail: Nikola@xray2.mskcc.org). Coordinates have been deposited in the Protein Data Bank (under accession codes 1FS1 (Skp1 Δ H8–F-box), 1FS2 (Skp1 Δ H8–F-box+LRRs) and 1FQV (Skp1–Skp2)).

Cultural references

The latest gadgets and media for cell and tissue culture.

Polystyrene microplates

Whatman www.whatman.com

Through a glass, darkly

Designed for use with optical measurements in cell culture analysis, these 96-well-format clear or black polystyrene microplates have a 0.175-mm-thick slip glass bottom and are designed to achieve a uniform level of background noise during optical analysis, irrespective of the position of the well within the plate. The black version is designed for use with sensitive probes, light-sensitive cells and background subtraction. Both clear and black versions are suitable for imaging studies with dry and oil-immersion optics, white light and laser excitation, conventional and confocal microscopy, single- and dual-wavelength probes and FRET and GFT techniques.

Reader Service No. 100



Whatman microplates, in black and white.

difficult cells such as algae, plant cells and mammalian lines. The design of these shakers is said to achieve better mixing and higher oxygen transfer rates at slower speeds than conventional shakers.

Reader Service No. 102

Cell culture brochure

NAPCO www.napco2.com

Everything but the kitchen sink

NAPCO's latest brochure covers the 'Cell Culture Specialist' range. Included are the company's CO₂ incubators, class II biosafety cabinets, ultra-low-temperature freezers and benchtop centrifuges. Specification tables are provided. In addition the company home page features a simple-to-use search facility for quick reference.

Reader Service No. 101

Infinity shaker

Hyperion Research www.hyperionresearch.com

Put a slow spin on it

Both the original and ultra-low-speed versions of this shaker feature continuously alternating motion, brushless motors, a universal power supply and automatic restart. The original model runs at speeds of 30–300 r.p.m. and can achieve up to 400 r.p.m. with a reduced load. The low-speed LS runs from 100 r.p.m. down to 3 r.p.m. and is intended for culturing



Mixing it: Hyperion's Infinity shaker.

Mesenchymal stem cells

BioWhittaker www.biowhittaker.com

Multipotent, self-renewing cells

Normal human mesenchymal stem cells will differentiate into multiple lineages in a process similar to that of normal tissue formation. Once committed to a lineage, the cells have the capacity to form several types of connective tissue, such that, for example, adipogenic lineage forms adipose tissue, chondrogenic lineage forms cartilage and osteogenic lineage forms bone. These cells are available cryopreserved in passage two. Around 750,000 cells are recoverable from each ampoule. They are positive for SH2, SH3, CD29 and CD44, and negative for CD14 and CD34. The cells also test negative for HIV-1, hepatitis-B and -C, mycoplasma, bacteria, yeast and fungi.

Reader Service No. 103

Titan microtitre plates

Radleys Discovery Technologies

www.radleys.com

Solving sticky problems

These plates are claimed to offer several advantages over polypropylene microtitre plates. They are chemically inert, do not leach extractables, can be used from cryogenic temperatures to +300°C, are less expensive than glass, do not fracture, and are non-stick. They are especially suitable for use with sticky samples or cells and the range includes 8-, 12-, 24-, 48-, 96- and 384-well plates and lids. Suggested applications include compound storage, combi-chem and high-throughput screening, as well as many biological assays and tissue culture applications.

Reader Service No. 104

BD 24-Multiwell Insert Systems

BD Bioscience www.bd.com

Automate chemotaxis and cell migration assays

BD 24-well Multiwell inserts are pre-coated with fibrillar collagen, collagen I or fibronectin to facilitate cell attachment, growth or differentiation. They are compatible with most robots and automated fluid handlers. All 24 inserts are part of a single unit and can be handled simultaneously. They are available as 3.0-µm pore size PET membrane inserts. Applications include real-time chemotaxis and cell migration assays.

Reader Service No. 105

Roller bottles

Nalge Nunc www.nalgenunc.com

Unbreakable bottles for adherent cell culture

Polyethylene terephthalate glycol copolymer is the material used to manufacture these bottles, which are 'virtually unbreakable' in normal use. The standard bottle comes in 1,050-cm² and 1,800-cm² sizes; the former will hold nearly 800 ml of suspension without fluid filling the neck when the bottle is horizontal. Expanded surface bottles have the same outer dimensions but use a pleated internal design that doubles the culture surface area.

Reader Service No. 106

HeLa cells

4C www.4c.be

They're only human

The 4C company, set up in 1989 to develop mammalian cell culture production technologies, produces HeLa cells as frozen whole-cell pellet, nuclear extract, nuclei, mitochondria, cytosol or purified protein. This HeLa line is intended for use in studies of biochemical processes or isolating cellular components of human origin, including splicing studies and high-throughput assays.

Reader Service No. 107



4C HeLa cells on ice.

new on the market

Sorvall Centritech

Kendro

www.kendro.com

Cell-processing system

Centritech systems are designed to gently separate animal cells in laboratory, pilot and production bioreactors. The design is based on the 'inverted question mark', in which one end of a flexible tube is rotated while the other end remains stationary, thus eliminating rotating seals. Fluid from the bioreactor is pumped peristaltically into the system, where viable cells are separated from dead cells and clarified supernatant. Live cells are then harvested or recycled to the bioreactor. By adjusting flow rates and centrifugal force, a portion of the dead cells and debris can be removed from the cell mass, further increasing productivity.

Reader Service No. 108

CollagenCult

StemCell

www.stemcell.com

Stain and archive haematopoietic colony assays

CollagenCult collagen-based media are optimized for the growth of human or murine haematopoietic colonies (BFU-E, CFU-GM and CFU-GEMM). Unlike methylcellulose-based media, CollagenCult allows the colony assays to be fixed and stained with antibodies against specific lineage markers, enabling definitive identification of different colony types that may otherwise be difficult to distinguish. The fixed and stained slides can be kept indefinitely to provide a permanent record of the assays.

Reader Service No. 109

XL-2020

Misonix

www.misonix.com

A sound investment

The XL-2020 ultrasonic liquid processor features a tactile keypad, programmable memory, LCD screen with user-friendly prompts, elapsed time and power-output displays, pulsing and timing functions, remote on/off capabilities and automatic overload protection.



Liquid engineering: the Misonix XL-2020.

Sonication aids in lysing of bacteria, virus, yeast and tissue cells in biological applications, and ultrasonic cavitation accelerates physical and chemical reactions used in the chemical industry. Other applications include environmental, industrial and pharmaceutical research and development.

Reader Service No. 110

TransferMan NK

Eppendorf

www.eppendorf.com

Electronic micromanipulator

This device combines intuitive control of movements via proportional kinetics with the advantages of electronic control, including storage of positions and selectable joystick transmission. Suggested applications include assisted reproduction through intracytoplasmic sperm injection, transfer of embryonic stem cells, micro-injection into suspension cells or cloning techniques. TransferMan can also be used for microdissection of single cells or cell groups from fixed preparations in pathology and cytology.

Reader Service No. 111

NE-PER Kit

Pierce

www.piercenet.com

Nuclear and cytoplasmic extraction reagents kit

The NE-PER kit keeps nuclear and cytoplasmic fractions separate and intact, and performs extractions from the same set of cultured cells or tissue in under two hours. The kit includes two cytoplasmic extraction

reagents and one nuclear extraction reagent. It is recommended for downstream applications including western blot analysis and gel shift, polymerase activity and enzyme activity assays.

Reader Service No. 112

Human cell cultures

TCS Biologicals

www.tcsgroup.co.uk

Ready-to-use human cell cultures

TCS Biologicals offer a range of normal human cells and media as ready-to-use proliferating cultures in either multiwell plate format or standard tissue-culture flasks. Large-vessel and microvessel endothelial cells, smooth muscle cells, keratinocytes, other epithelial cells and fibroblasts are available. Customized cell-based assays can be performed and post-treatment samples such as cell pellets, lysates, nucleic acid, protein or nuclear extracts supplied.

Reader Service No. 113

Cellhouse 170 HI

Heto-Holten A/S

www.heto-holten.com

No fans for this CO₂ incubator

Heto-Holten, now part of the Jouan group, specializes in biotechnology production equipment. The Cellhouse 170 HI incubator includes a range of features for accurate performance and control. The incubator uses the latest electronics for the precise measurement of temperature and CO₂ levels. The incubator offers operators a proactive control system, which continuously surveys temperature, CO₂ and humidity and reacts automatically to optimize conditions within the chamber. A heat-profiling system eliminates the need for a fan and, as a result, the potential for contamination is dramatically reduced whilst culture conditions are improved.

Reader Service No. 114

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

www.alphainnotech.com

Alpha Innotech Corporation

READER ENQUIRY NO. 48

TREVIGEN

new PARP

Poly (ADP-ribose) Polymerase Reagents

- PARP inhibition assay
- Biotinylated NAD
- Recombinant human PARP
- Anti PARP and PAR

www.trevigen.com

US and Canada: 1 800 873 8443
TEL: 301 216 2800 • FAX: 301 216 2801

READER ENQUIRY NO. 50

Look no further

for the best value in Oligo Synthesis

Competitive pricing
Routine 24 hour dispatch • Every Oligo QC verified by PAGE
Rigorous batch QC by PCR and Mass Spec
Web ordering and tracking • Local Freephone numbers
A great deal in any language!!

SIGMA GENOSYS

(+44) (0) 1223 839000 • fax (+44) (0) 1223 839200
email: info@sigma-genosys.co.uk • www.sigma-genosys.co.uk

READER ENQUIRY NO. 33